OPTIMAL CONTROL OF CONTINUOUS AND DISCONTINUOUS FLOW

Name: CRISTIAN A. HOMESCU
Department: Department of Mathematics
Major Professor: I.M. Navon
Degree: Doctor of Philosophy
Term Degree Awarded: Summer, 2002

Numerical and theoretical aspects of solving optimal control problems for a continuous flow (*suppression of the Karman vortex street for a flow around a cylinder*) and for a discontinuous flow (*changing the location of discontinuities for the shock-tube problem*) are considered.

The minimization algorithms require the gradient (or a subgradient) for the smooth (respectively non smooth) cost functional. The numerical value of the gradient (respectively a subgradient) is obtained using the adjoint method.

The optimal solutions are verified using their physical interpretation. A very convincing argument for the validity of the numerical optimal solutions is obtained comparing the values corresponding to observed physical phenomena to the above-mentioned numerical optimal controls.

Sensitivity analysis of a discontinuous flow, namely for the shock-tube problem of gas dynamics, was also studied. Better results are obtained compared to the available literature, due to the use of adaptive mesh refinement.

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

OPTIMAL CONTROL OF CONTINUOUS AND DISCONTINUOUS FLOW

By

CRISTIAN A. HOMESCU

A dissertation submitted to the
Department of Mathematics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Summer Semester, 2002

The members of the Committee approve the dissertation of CRISTIAN A. HOMESCU defended on July 15, 2002.

I.M. Navon
Professor Directing Thesis

R. Pfeffer
Outside Committee Member

M.Y. Hussaini
Committee Member

G. Erlebacher
Committee Member

S. Blumsack
Committee Member

Approved:

DeWitt Sumners, Chair
Department of Mathematics

To my family: my mother, my father and my sister…

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to Prof I.M. Navon, who directed this research. First by choosing a topic which proved to be very interesting and challenging as well as applicable in many domains, then by constant encouragements and advice, Prof. Navon was a perfect mentor. His expertise in the field proved to be extremely valuable for the completion of my research.

I am also very grateful to Prof. M.Y. Hussaini for his many insightful comments. Prof. Hussaini encouraged me constantly to look beyond the mathematical results and understand the physical phenomena corresponding to these results.

I would like to thank Prof. G. Erlebacher for finding the time to help me understand many numerical and flow visualization subtleties.

I would like to thank also Prof. S. Blumsack and Prof. R. Pfeffer for being members of my PH.D committee. It is an honor to present my research in front of such accomplished scientists.

The support of the Department of Mathematics and of the School of Computational Science and Information Technology at Florida State University is gratefully acknowledged.

Finally I would like to thank all my friends here in Tallahassee, too many to enumerate, who made my stay at F.S.U. a very enjoyable experience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Numerical and theoretical aspects of solving optimal control problems for a continuous flow (*suppression of the Karman vortex street for a flow around a cylinder*) and for a discontinuous flow (*changing the location of discontinuities for the shock-tube problem*) are considered.

The minimization algorithms require the gradient (or a subgradient) for the smooth (respectively non smooth) cost functional. The numerical value of the gradient (respectively a subgradient) is obtained using the adjoint method.

The optimal solutions are verified using their physical interpretation. A very convincing argument for the validity of the numerical optimal solutions is obtained comparing the values corresponding to observed physical phenomena to the above-mentioned numerical optimal controls.

Sensitivity analysis of a discontinuous flow, namely for the shock-tube problem of gas dynamics, was also studied. Better results are obtained compared to the available literature, due to the use of adaptive mesh refinement.

CHAPTER 1

INTRODUCTION

Real-world applications arising from very different fields: fluid dynamics (Sritharan [179], Gunzburger [91]), engineering (Siouris [176]), mechanics (Akulenko [3]), credit risk (Cossin [40]), management science and economics (Sethi [173], Seierstad [172]), thermodynamics (Berry [15]), chemistry (Edgar [50]), biomedicine (Swan [181]), electric power systems (Christensen [33]), distributed nuclear reactors (Christensen [34]), hydrosystems (Mays [142]) can be formulated as optimal control problems following a general description:

Influence the behavior of the system
so as to achieve a desired goal.

This is the equivalent to controlling the system by selecting a certain set of the parameters that determine its behavior. The *optimal* parameters are obtained by performing the minimization of a given cost functional measuring the discrepancy between model and observations in a given time interval.

The characteristics of the cost functional determine which optimization method is better suited for solving the minimization problem. For the subset of differentiable cost functions smooth optimization methods are more efficient, while non smooth optimization algorithms are more appropriate for the subset of non differentiable cost functionals. The object of this dissertation is to provide both a theoretical analysis as well as a numerical solution for optimal control problems representative of each category.

An optimal control problem for a viscous flow past a circular cylinder is chosen for the case of a differentiable cost functional. Optimal control for the shock-tube problem is considered for the case of a non smooth cost functional. Sensitivity analysis for the discontinuous flow case is also studied.

1.1 Optimal control for flow past a circular cylinder

The viscous flow past a circular cylinder has been extensively studied due to its simple geometry and its representative behavior of general bluff body wakes. A deep understanding of the control strategies necessary to control flows past rotating bluff bodies could be applied in areas like drag reduction, lift enhancement, noise and vibration control, aerodynamics etc.

A very important characteristic of this flow is the Karman vortex shedding (which has been extensively studied for the last 90 years, starting with the pioneering work of Von Karman [119]).

Research on the problem of a flow past a cylindrical rotating body has been the subject of many experimental (Badr et al. [8], [9], Tokumaru and Dimotakis [188]), and numerical investigations (Chen et al. [28], Baek and Sung [10], Dennis et al. [46], Juarez et al. [117], Chou [31]). However most of these results are primarily focused on the study of formation and development of vortices in a cylinder wake and they do not attempt to suppress vortex shedding.

Examples of applying active control of vortex shedding in experiments are given by Gad-el-Hak [60], [61] and Modi [144]. Modi's experiments are related to the moving surface boundary layer control for airfoils. The moving surfaces are provided by rotating cylinders located at the leading edge and/or trailing edge as well as the top surface of an airfoil. It has been shown that this mechanism of moving surfaces can prevent flow separation by retarding the initial growth of the boundary layer, with important consequences for lift enhancement and stall delay. The control parameter used was the speed ratio (which represents the ratio of cylinder speed to the free stream speed). This speed ratio can be either constant in time or time-dependent (e.g., if the airfoil is undergoing a rapid maneuver). This type of result provided us with the motivation to consider flow control for either a constant or time-dependent angular rotation of the cylinder.

Different approaches for the control of a flow around a cylinder have been successfully employed in the last two decades. For example, Tang and Aubry [184] suppressed vortex shedding by inserting two small vortex perturbations in the flow; Gillies [68] used neural networks; Gunzburger and Lee [89] determined the amount of fluid injected or sucked on rear of the cylinder from a feedback law depending on pressure measurements at stations along the surface of the cylinder; Huang [106] suppressed vortex shedding by feedback sound; Joslin et al. [116] showed that flow instabilities can be controlled by wave cancellation; Kwon and Choi [123], Ozono [156] and You et al. [207] employed splitter plates attached to the cylinder; Park et al. [157] used a pair of blowing/suction slots located on the surface of the cylinder; Sakamoto and Haniu [167] introduced a smaller cylinder near the main cylinder, with experiments conducted by changing the gap between the cylinders and the angle along circumference from the front stagnation point of the main cylinder; the flow is controlled via cylinder rotation (e.g., Tang et al. [183], Tao et al. [185], Warui and Fujisawa [203], He et al. [98], or Tokumaru and Dimotakis [189]); Pentek and Kadtke [159] implemented a chaos control scheme to capture and stabilize a concentrated vortex around the cylinder, the control being actuated by uniformly rotating the cylinder and actively changing the background flow velocity far from the body.

Due to the complexity and large dimensions of the control problem, suboptimal control strategies have been considered and implemented. The concept of *instantaneous control* (e.g., control at every time step of the underlying dynamical systems) was applied in Choi et al. [30]. Another approach involves two stages:*first* the approximation of the equations of the fluid flow using reduced order models and *second* an exact optimization for the reduced system, the difference among various research efforts consisting in the choice of the basis functions used for the reduced models. In the *reduced basis* approach one uses as basis functions the terms which arise in series expansion of the solution with respect to a parameter (e.g., Ito and Ravindran [112]). The *proper orthogonal decomposition* (POD) approach is applied by Graham et al. [77], [78] and Afanasiev and Hinze [2] .

Optimal control methods (**OCM**) have been employed for flow control. *Distributed controls* were used by Abergel and Temam [1], Gunzburger et al. [88], Hou et al. [105], [104]; *blowing and suction* on the surface of the cylinder was studied by Berggren [14], Bewley [17], Ghattas and Bark [63], Li et al. [131]; *velocity tracking* (boundary velocity controls) was employed by Gunzburger and Manservisi [90], Gunzburger et al. [87], Hou and Ravindran [103], [102].

A key component of the process of flow control is the minimization of a cost functional aiming at the optimization of some of the flow characteristics.

Abergel and Temam [1] minimized the turbulence for a flow respectively driven by volume forces, a gradient of temperature and a gradient of pressure (the turbulence being measured by a L^2 norm of the curl of v ($\|\nabla \times v\|_{L^2}$) or, alternatively, by studying the stress at the boundary); Berggren [14] minimized the vorticity field. Bewley et al. [17] reduced the turbulent kinetic energy and drag; Ghattas and Bark [63] used as objective function the rate at which energy is dissipated in the fluid.

Our research presents the numerical solution to the problem of controlling vortex shedding for a flow past a rotating cylinder using optimal control methods. It is shown that the nature of the vortex shedding process is significantly altered by cylinder rotation. We employ a global control approach (the entire body is subjected to prescribed motion), as compared to a local control method (e.g., blowing/suction as reported by Li et al. [131]).

The mathematical formulation of the problem implies minimization of a cost functional. Since all efficient local minimization algorithms require the computation of the gradient of an objective functional (functional described in chapter 7) with respect to the control parameters, part of this effort was dedicated to the gradient computation.

The adjoint method was employed to obtain the gradient of the discrete cost functional. The adjoint was constructed directly from the source code of the original discrete nonlinear model, circumventing difficulties which would appear if one were to first obtain the continuous adjoint model and then discretize the adjoint equations (the differences between the *differentiate-then-discretize* approach and the *discretize-then-differentiate* approach are discussed by Gunzburger [86]).

The objective functional included a regularization term since preliminary numerical results suggested ill-posedness of the optimization problem. We chose the regularization term to be from the class of Tikhonov regularization (Tikhonov and Arsenin [187]). Another important characteristic is the length of the "control" window (the time window employed for minimization). It was found that the length of this time window should be larger than the vortex shedding period if the angular velocity (which serves as the control parameter) is time-dependent. However, if the angular velocity is constant in time, the length of the time window should only exceed a certain threshold value which can be smaller than the vortex shedding period.

The results obtained show that vortex shedding is suppressed for the Reynolds number in the range: $40 \leq Re \leq 1000$. The regimes of flow change for different subsets of the range considered. The flow characteristics are different for $40 \leq Re \leq 150$, $150 \leq Re \leq 250$ and $250 \leq Re \leq 1000$ respectively. For the same values of optimal rotation rate employed to achieve the elimination of the vortex shedding, the time histories of the drag coefficient show that a significant reduction in the amplitude of its variation is obtained compared to the case of the fixed cylinder.

As far as we know our research is the first apply numerical optimal control methods for the flow control problem around a rotating cylinder. Our method converged for both cases considered: constant rotation in time or time-dependent rotation.

Comparable results were obtained for constant rotation (Kang et al. [118], Chew et al. [29], Badr et al. [9] and Chou [32]) and, respectively, for the time-dependent rotation (Tokumaru and Dimotakis [188], Baek and Sung [10] and He et al. [98]). The main difference between our approach and their research is the following: they obtained the values of the rotation parameters for which the flow has the required characteristics by experiments or active control applied to numerical simulations.

1.2 Sensitivities for a flow with discontinuities

Sensitivities (for both continuous and discontinuous flows) are derivatives of the variables or cost functionals that describe the model with respect to parameters that determine the behavior of the model (e.g., initial conditions, boundary conditions, shape parameters). They provide information about what, where and when these parameters most influence the model output. Employed in an optimization setting they help determine the gradient of the objective functional used in the optimization process.

The sensitivity analysis (**SA**) means very different things to different people (compare the reviews of Turanyi [191], Janssen et al. [115], Helton [99] and Goldsmith [72]) but all its applications share a common goal: *to investigate how a given computational model responds to variations in its inputs.*

We studied **SA** for a fluid dynamics problem (characterized by several types of discontinuities). Besides fluid dynamics, **SA** proved to be very useful in many other scientific fields. To exemplify the extent of **SA** applications we mention a very recent **SA** handbook by Satelli et al. [170] which describes the principles of sensitivity analysis in various settings and presents many **SA** methodologies for ecology, chemistry, mechanics, economics and policy-making, to mention but a few.

The vast majority of **SA** applications were obtained for problems involving continuous functions. Research was also performed in the presence of discontinuities but many questions in this area remain yet unanswered. Sensitivity analysis in the case of a model with discontinuities was applied in fluid dynamics, aerodynamics, chemistry, financial analysis, meteorology or environmental studies, and the list goes on.

Discontinuous **SA** studies include shape optimization for fluids (Burgreen and Baysal [22], Newman et al. [149], Taylor et al. [186], Mohammadi and Pironneau [145]), noise analysis and optimization of electronic circuits (Nguyen et al. [150]), control of contaminant releases in rivers (Piasecki and Katopodes [160]), control of water movement through systems of irrigation canals (Sanders and Katopodes [168]), shallow water wave control (Sanders and Katopodes [169]), aeroelastic analysis (Giunta and Sobiesczanski-Sobieski [69]), shock sensitivity evaluations of dynamic financial strategies (Gourieroux and Jasiak [75]) and meteorological applications (Zhang et al. [212]).

Theoretical and computational aspects of sensitivity calculation in the presence of discontinuities were also presented by Ulbrich [193], Cliff et al. [36], Godlewski and Raviart [71], Bouchut and James [21] and DiCesare and Pironneau [47].

Numerical sensitivities were computed by Narducci et al. [147] for optimization of duct flow with a shock using quasi-one-dimensional Euler equations. In their research they employed continuous (*differentiate-then-discretize*) and discrete (*discretize-then-differentiate*) methods to compute the design sensitivities. The continuous method requires analytical expressions for the derivatives of the velocity and shock location with respect to the design variables derived from the governing equations and the shock jump conditions (the difference between direct and adjoint method in this case is that the adjoint method avoids computing these derivatives directly). For the discrete method a coordinate-straining approach with a shock penalty was employed (to avoid difficulties caused by the presence of non smooth functions).

For the same problem as Narducci et al. (quasi-one dimensional duct flow) Cliff et al. [36] introduced the shock location as an explicit variable which allowed one to fit the shock and yielded a problem with sufficiently smoothed functions.

Cliff et al. [37] carried out sensitivity calculations for the 1-D Euler system. No numerical calculations were performed however.

Our research is focused on the numerical computation of flow sensitivities with respect to an initial flow parameter for the shock-tube problem (1-D Riemann problem for the Euler equations) for which the exact values of the flow sensitivities are known.

We chose the discrete (*discretize-then-differentiate*) approach which in our opinion is more suitable than the continuous approach for flows with discontinuities. Our numerical results were compared to the results presented by Gunzburger [86] and they proved to solve better the regions with discontinuities due to the use of adaptive mesh refinement.

1.3 Optimal control of the 1-D Riemann problem of gas dynamics

Recently optimal control involving non smooth functions has attracted the attention of an increasing number of researchers due to availability of new methods of non differentiable optimization employing subgradients following the seminal work of Lemarechal [125] (e.g., Lemarechal [126], Bonnans et al. [20], Schramm and Zowe [171], Luksan and Vlcek [136], Makela and Neittaanmaki [138] to cite but a few).

Non smooth cost functionals were employed in variational data assimilation in atmospheric sciences (Zhang et al. [212]), for inverse design problems involving transonic diffusers : 1-D (Narducci et al. [147]) or 2-D (Dadone et al. [44]), in acoustics (Habbal [93]), for the research of a convex hull with bounded curvature of a given set of points (Hassold [97]), in mechanical structures (minimizing the maximal stress over an arch structure Habbal [92]), for chromatography (James and Sepulveda [113]), capital asset management (Leonard and Long [127]), in the design of a duct flow with a shock (Frank and Shubin [58], Cliff et al. [36], Iollo et al. [111]), for airfoil design (Jameson [114], Matsuzawa and Hafez [140], [141], Iollo and Salas [110] and Giles and Pierce [66]).

The presence of discontinuities creates serious theoretical and numerical difficulties. Good shock-capturing schemes with low continuity properties often cannot be combined successfully with efficient optimization methods requiring smooth functions (e.g., gradient-based methods). To alleviate this problem one can use methods that are relatively insensitive to the non smoothness of the cost function. Stochastic optimization methods were applied for the design of a minimum time changeover operation for a pressure vessel avoiding the formation of explosive mixtures (Barton et al. [12]) or for aerodynamic shape optimization (Huyse and Lewis [108]). Genetic algorithms (Oyama et al. [155]) were also used for wing optimization. For these non-gradient-based methods the drawback is the very large number of analyses required (i.e., large memory demands) as the number of variables increases.

In the case of gradient-based methods different remedies to alleviate the influence of the discontinuities were employed. For variables continuous across the shock one can avoid dealing with shocks by considering cost functions based on the above variables (e.g., the surface flux for inverse nozzle design as used by Matsuzawa and Hafez [140]). For most cases the shocks were smoothed using numerical dissipation. It was shown that sometimes smoothing is equivalent to modifying the cost function (Matsuzawa and Hafez [140]). An alternative smoothing procedure has been introduced by Valorani and Dadone [197], namely a filtering process which was obtained by modifying a set of sensitivity equations by adding artificial dissipative terms. The optimization search was performed on the original non smooth objective function computed with an accurate (non smoothed) flow analysis but with smoothed flow sensitivities.

If the shocks are weak at design conditions (e.g., transonic flows) acceptable results can be obtained by addition of artificial dissipation. However, accurate treatment of the shock waves is essential in other cases (e.g., supersonic flows). The alternative approach to shock smearing is shock fitting which involves careful integration of the objective function through the shock wave (Narducci et al. [147]). Perturbation of a discontinuous function produces delta functions and formulations based on variations of smooth functions have to be modified (Iollo et al. [111]). Another approach was to introduce the shock location as an explicit control variable (Cliff et al. [36]). A coordinate straining method was also employed by Narducci et al. [147]. It consists of a coordinate transformation aimed at aligning the calculated shock with the target, followed by addition of a penalty term proportional to the square of distance between the shocks.

Results for the optimal control of the Euler equations were obtained, among others, by Anderson and Venkatakrishnan [6] (in 2-D), Arian and Salas [7] (in 2-D), Dadone and Grossman [43], [42] (2-D and 3-D), Cliff et al. [35], [36], [37], [38]) (1-D and 2-D).

Theoretical contributions (combined with practical applications in certain cases) for the adjoint method were provided by Giles and Pierce [64], [65], [66], [67] (for Euler equations) and Ulbrich [192], [193], [195], [194] (in the setting of optimal control for scalar conservation laws). A generalized adjoint for physical processes in atmospheric sciences with parameterized discontinuities was studied by Xu [206]. Numerical aspects of the adjoint model for discontinuous nonlinear atmospheric models were discussed by Zhang et al. [211].

Problems with discontinuities in an optimal control setting or in sensitivity-based control were studied by Mohammadi and Pironneau [145], Gunzburger [86], [85], Tolsma and Barton [190] and Zhang et al. [212].

Practical aspects of control of problems with shocks were presented by Iollo and Salas [109], Birkemeyer et al. [18], Stanewsky [180], Jameson [114], Bein et al. [13] or Wang et al. [202].

Our research consists of theoretical and numerical results for an optimal control problem of the unsteady 1-D Riemann problem of Euler equations (shock-tube). The numerical solutions of the optimal control problem were obtained using both non smooth and smooth optimization algorithms.

This specific problem was chosen due to the fact that it has an analytical solution which is characterized by the presence of several types of discontinuities: shocks, contact discontinuities and wave rarefaction regions. This Riemann problem may be briefly described in the following way: a gas tube is divided by a membrane into two regions with different values for pressure and density fields and a zero velocity field. After the membrane is suddenly removed the gas moves freely.

Our optimal control problem has very interesting aerodynamic applications, consisting in moving the regions of discontinuities to *desired* locations by matching the *desired* flow to the numerical flow. The control parameters consist of the initial values of pressure and density to the left and to the right of the membrane. We consider the initial velocity to the left and to the right of the membrane to be zero. The cost functional is the weighted L^2 difference between the *observations* and the numerical values for density, pressure and velocity fields. The observations are computed from the analytical solution of the Riemann problem in two ways: either at the end of the assimilation window or distributed in time within the assimilation window.

Two numerical models were chosen, representative of possible approaches for solving a flow with discontinuities: a high-resolution model (**HRM**) and a model with artificial viscosity (**AVM**).

We employed a non smooth optimization algorithm (**PVAR**), developed by Luksan and Vlcek [136], [137], [198]. We also used a smooth optimization algorithm (**L-BFGS**), described in Nocedal [151] and Liu and Nocedal [134]). Both methods require the computation of a subgradient (respectively the gradient) of the cost functional. This subgradient (respectively gradient) is obtained from the adjoint model derived from the original numerical model. Accuracy tests for both the gradient and subgradient obtained via the adjoint method are presented.

We considered two time horizons which are representative for the time evolution of the flow. Their length was chosen for two main reasons. First we wanted to ensure that all desired characteristics of the discontinuities are still present in the flow at the end of each time window. Second, we selected the larger time window such that if we were to slightly increase it some of the discontinuity characteristics will disappear from the spatial domain considered.

We obtained excellent results using non smooth optimization for both models and for both time horizons. The numerical flow corresponding to the *optimized* initial conditions matches closely the observations and the location of the discontinuities was changed to the *desired* location. The figures describing the evolution of entropy at various stages of the minimization process show that the numerical solution satisfies the entropy condition which is a requirement for a physical solution of the shock-tube problem.

The **L-BFGS** algorithm did not converge in many cases. Even for the cases where convergence was obtained one may notice a large difference between the **L-BFGS** optimization results and the *desired* values of the control parameters.

For the model with artificial viscosity a discontinuity detection method was used to eliminate the points where the shock is located from the computation of the cost functional and its gradient (or subgradient). As a result, the optimized results were obtained at the same level of accuracy but in fewer minimization iterations.

CHAPTER 2

THEORETICAL FRAMEWORK FOR OPTIMAL CONTROL AND SENSITIVITY ANALYSIS

2.1 General characteristics of an optimal control problem

Every problem of optimal control is characterized by several main features.

It has an **objective**, i.e. a reason why one wants to control the system. There are numerous objectives of interest in applications, e.g., drag minimization, lift enhancement, preventing transition to turbulence, reducing noise, personnel task scheduling, shape optimization, control of heat transfer, operation of a cascade of power stations, mineral resource extraction in an open economy, stock selections. Mathematically, such an objective is expressed as a cost functional.

Constraints must be imposed on candidate optimizers. The constraints are derived from the given law according to which the system evolves. They are expressed in terms of a specific set of equations. One may mention here partial differential equations **PDE** (e.g., Navier-Stokes or Euler equations for incompressible or compressible flows, heat equation, shallow-water equations, Black-Scholes equations for financial mathematics), ordinary differential equations **ODE** (chemical reactions, spreading of diseases), stochastic differential equations **SDE** (noisy evolution of stock values or porous media flow) and differential-algebraic equations **DAE** (for dynamical models).

The nature of the state equations and of the boundary conditions is determined by the mathematical model adopted. For this model one can identify a group of dependent variables called **state variables** (e.g., velocity, pressure, density, temperature, energy).

Finally one has **control parameters** which determine the behavior of the system. For the fluid applications one can have *boundary value controls* (injection or suction, heating or cooling), *distributed controls* (heat sources or magnetic fields) or *shape controls* (exit area for a nozzle, movable walls, leading or trailing edge flaps).

The optimal control problem ($\mathcal{OCP}$) is then stated as:

Find controls g and states Φ such that the cost functional $\boldsymbol{\mathcal{J}}(\Phi, g)$ is minimized
subject to the flow equations $\mathcal{FLOW}(\Phi, g) = 0$ ($\mathcal{OCP}$)

The set of admissible controls is the set of all controls g allowed by the physical limitation of the problem. The optimal control g^* which solves ($\mathcal{OCP}$) is selected from the set of admissible controls denoted by $\mathcal{U}_{ad}$.

2.2 The adjoint approach for solving an optimal control problem

We follow the adjoint approach as introduced by Talagrand and Courtier ([182]).

The following two basic properties of Hilbert spaces form the basis of this approach.

If $\mathcal{B}$ is a Hilbert space with inner product denoted by $(,)$ and $\mathbf{v} \rightarrow \mathbf{F}(\mathbf{v})$ a differentiable scalar function defined on $\mathcal{B}$, then the differential of $\mathbf{F}$ can be expressed as

$$\delta \mathbf{F} = (\nabla_{\mathbf{v}} \mathbf{F}, \delta \mathbf{v}) \tag{2.1}$$

where $\nabla_{\mathbf{v}} \mathbf{F}$ is the "gradient" of $\mathbf{F}$ with respect to $\mathbf{v}$.

Let $\mathcal{C}$ be another Hilbert space with inner product denoted by $<,>$ and $\mathbf{L}$ a continuous linear operator from $\mathcal{B}$ to $\mathcal{C}$. There exists a unique continuous linear operator $\mathbf{L}^*$, called the *adjoint operator* of $\mathbf{L}$, from $\mathcal{C}$ to $\mathcal{B}$ such that

$$(\mathbf{v}, \mathbf{L}\mathbf{z}) =< \mathbf{L}^* \mathbf{v}, \mathbf{z} > \tag{2.2}$$

for any $\mathbf{v} \in \mathcal{C}$ and any $\mathbf{z} \in \mathcal{B}$.

Consider a differentiable function $\mathbf{z} \rightarrow \mathbf{v} = \mathbf{G}(\mathbf{z})$ of $\mathcal{B}$ into $\mathcal{C}$. The function $\mathbf{F}$ is a composite function of $\mathbf{z}$ ($\mathbf{F}(\mathbf{v}) = \mathbf{F}[\mathbf{G}(\mathbf{z})]$). Then the differential of $\mathbf{v}$ is equal to

$$\delta \mathbf{v} = \mathbf{G}^{'} \delta \mathbf{z} \tag{2.3}$$

where $\mathbf{G}^{'}$ is the linear operator obtained by differentiation of $\mathbf{G}$.

Introducing the adjoint $\mathbf{G}^{'*}$ of $\mathbf{G}^{'}$ and using (2.3) one obtains

$$\delta \mathbf{F} = (\nabla_{\mathbf{v}} \mathbf{F}, \mathbf{G}^{'} \delta \mathbf{z}) =< \mathbf{G}^{'*} \nabla_{\mathbf{v}} \mathbf{F}, \delta \mathbf{z} > \tag{2.4}$$

This shows that the gradient $\nabla_{\mathbf{u}} \mathbf{F}$ of $\mathbf{F}$ with respect to $\mathbf{z}$ is equal to

$$\nabla_{\mathbf{z}} \mathbf{F} = \mathbf{G}^{'*} \nabla_{\mathbf{v}} \mathbf{F} \tag{2.5}$$

The formula (2.5) is at the basis of the use of adjoint equations in control theory. Assuming that the operation $\mathbf{z} \rightarrow \mathbf{v} = \mathbf{G}(\mathbf{z})$ denotes the integration of the numerical model, the formula (2.5) provides a very efficient way for the numerical computation of the gradient $\nabla_{\mathbf{z}} \mathbf{F}$.

We present now details of the numerical computation of the gradient based on the above discussion. Let us assume that the model evolution equation is written as

$$\begin{aligned} \frac{d\mathbf{U}(\mathbf{X}, \mathbf{Y}, t)}{dt} &= \mathbf{F}(\mathbf{U}, \mathbf{Y}, t) \\ \mathbf{U}(t_0) &= \mathbf{U}_0 \end{aligned} \tag{2.6}$$

where $\mathbf{X} = (X_1, \ldots, X_m) \subset \mathbb{R}^m$ is the position vector, $\mathbf{U}(\mathbf{X}) = [U_1(\mathbf{X}), \ldots, U_K(\mathbf{X})]$ is the state vector which belongs to a Hilbert space whose inner product is denoted by $<,>$ and $\mathbf{Y}(\mathbf{X}) = [Y_1(\mathbf{X}), \ldots, Y_P(\mathbf{X})]$ the vector of system parameters.

We consider the cost functional

$$\boldsymbol{\mathcal{J}} = \int_{t_0}^{t_W} \mathbf{H}[\mathbf{U}, \mathbf{U}^{obs}, t]\, dt \tag{2.7}$$

where $[t_0, t_W]$ is the length of the assimilation window and $\mathbf{H}$ is a functional depending on the state vector $\mathbf{U}$ and the observations $\mathbf{U}^{obs}$ available at time t.

For a given initial condition $\mathbf{U}_0$ and a given vector of system parameters $\mathbf{Y}$ there exists a solution $\mathbf{U}(t)$ of (2.6). The first order variation $\delta\boldsymbol{\mathcal{J}}$ is equal to

$$\delta\boldsymbol{\mathcal{J}} = \int_{t_0}^{t_W} < \nabla_{\mathbf{U}}\mathbf{H}(t), \delta\mathbf{U}(t) > \, dt \tag{2.8}$$

where $\nabla_{\mathbf{U}}\mathbf{H}(t)$ is the gradient of $\mathbf{H}$ with respect to $\mathbf{U}$ taken at point $(\mathbf{U}(t), t)$ and $\delta\mathbf{U}(t)$ is the first-order variation of $\mathbf{U}(t)$ resulting from the perturbations $\delta\mathbf{U}_0$ and $\delta\mathbf{Y}$ of $\mathbf{U}_0$ and respectively $\mathbf{Y}$.

The variation $\delta\mathbf{U}(t)$ is obtained from $\delta\mathbf{U}_0$ and $\delta\mathbf{Y}$ by integrating the *tangent linear model* relative to the solution $\mathbf{U}$

$$\frac{d[\delta\mathbf{U}(t)]}{dt} = \mathbf{F}^{'}(t)\delta\mathbf{U} \tag{2.9}$$

where $\mathbf{F}^{'}$ is the operator obtained by differentiating $\mathbf{F}$ with respect to $\mathbf{U}$, taken at point $\mathbf{U}(t)$. The solution of the linear equation (2.9) can be written as

$$\delta\mathbf{U}(t) = \mathbf{R}(t, t_0)\delta\mathbf{U}_0 \tag{2.10}$$

where $\mathbf{R}(t, t_0)$ is a linear operator called the *resolvent* between times t and t_0.

Equation (2.8) can now be rewritten as

$$\begin{aligned}
\delta\boldsymbol{\mathcal{J}} &= \int_{t_0}^{t_W} < \nabla_{\mathbf{U}}\mathbf{H}(t), \mathbf{R}(t, t_0)\delta\mathbf{U}_0 > \, dt \\
&= \int_{t_0}^{t_W} < \mathbf{R}^*(t, t_0)\nabla_{\mathbf{U}}\mathbf{H}(t), \delta\mathbf{U}_0 > \, dt \\
&= \left\langle \int_{t_0}^{t_W} \mathbf{R}^*(t, t_0)\nabla_{\mathbf{U}}\mathbf{H}(t)\, dt, \delta\mathbf{U}_0 \right\rangle \qquad (2.11)
\end{aligned}$$

where $\mathbf{R}^*(t, t_0)$ is the adjoint of $\mathbf{R}(t, t_0)$.

We can see from (2.8) and (2.11) that the gradient of $\boldsymbol{\mathcal{J}}$ with respect to $\mathbf{U}_0$ is

$$\nabla_{\mathbf{U}_0}\boldsymbol{\mathcal{J}} = \int_{t_0}^{t_W} \mathbf{R}^*(t, t_0)\nabla_{\mathbf{U}}\mathbf{H}(t)\, dt \tag{2.12}$$

We introduce the adjoint equation of (2.9), using the adjoint vector $\delta^{'}\mathbf{U}(t)$ corresponding to $\delta\mathbf{U}(t)$:

$$-\frac{d[\delta^{'}\mathbf{U}(t)]}{dt} = \mathbf{F}^{'*}(t)\delta^{'}\mathbf{U} \tag{2.13}$$

where $\mathbf{F}^{'*}$ is the adjoint of $\mathbf{F}^{'}$.

We denote by $\mathbf{S}(t^{'},t)$ its resolvent between times t and $t^{'}$:

$$\delta^{'}\mathbf{U}(t) = \mathbf{S}(t^{'},t)\delta^{'}\mathbf{U}(t^{'}) \tag{2.14}$$

For any two solutions $\delta\mathbf{U}(t)$ and $\delta^{'}\mathbf{U}(t)$ of the direct and adjoint equations (2.9) and (2.13) respectively, the inner product $< \delta\mathbf{U}(t), \delta^{'}\mathbf{U}(t) >$ is constant in time since

$$\begin{aligned}\frac{d}{dt} < \delta\mathbf{U}(t), \delta^{'}\mathbf{U}(t) > &= \left\langle \frac{d\delta\mathbf{U}(t)}{dt}, \delta^{'}\mathbf{U}(t) \right\rangle + \left\langle \delta\mathbf{U}(t), \frac{d\delta^{'}\mathbf{U}(t)}{dt} \right\rangle \\ &=< \mathbf{F}^{'}(t)\delta\mathbf{U}(t), \delta^{'}\mathbf{U}(t) > - < \delta\mathbf{U}(t), \mathbf{F}^{'*}\delta^{'}\mathbf{U}(t) >= 0\end{aligned}$$

Let $\mathbf{Z}$ and $\mathbf{Z}^{'}$ be any two elements in the Hilbert space considered. The solution of the direct equation (2.9) defined by the initial condition $\mathbf{Z}$ at time t assumes at time $t^{'}$ the value $\mathbf{R}(t^{'},t)\mathbf{Z}$ while the solution of the adjoint equation (2.13) defined by the condition $\mathbf{Z}^{'}$ at time $t^{'}$ assumes at time t the value $\mathbf{S}(t,t^{'})\mathbf{Z}^{'}$. Therefore we have

$$< \mathbf{R}(t^{'},t)\mathbf{Z}, \mathbf{Z}^{'} >=< \mathbf{Z}, \mathbf{S}(t,t^{'})\mathbf{Z}^{'} > \tag{2.15}$$

The relation (2.15) is valid for any elements $\mathbf{Z}$ and $\mathbf{Z}^{'}$, which shows that $\mathbf{S}(t,t^{'})$is the adjoint of $\mathbf{R}(t^{'},t)$. In other words the resolvent of the adjoint equation between $t^{'}$ and t is the adjoint of the resolvent of the direct equation between t and $t^{'}$.

The expression (2.12) then becomes

$$\nabla_{\mathbf{U}_0}\boldsymbol{\mathcal{J}} = \int_{t_0}^{t_W} \mathbf{S}(t_0,t)\nabla_{\mathbf{U}}\mathbf{H}(t)\,dt \tag{2.16}$$

We consider next the “inhomogeneous adjoint equation”:

$$-\frac{d\delta^{'}\mathbf{U}}{dt} = \mathbf{F}^{'*}(t)\delta^{'}\mathbf{U} + \nabla_{\mathbf{U}}\mathbf{H}(t) \tag{2.17}$$

with initial condition:

$$\delta^{'}\mathbf{U}(t_W) = 0 \tag{2.18}$$

The solution of (2.17)-(2.18) is

$$\delta^{'}\mathbf{U}(t) = \int_{t}^{t_W} \mathbf{S}(t,\tau)\nabla_{\mathbf{U}}\mathbf{H}(\tau)\,d\tau \tag{2.19}$$

Comparing (2.16) and (2.19) we can see that

$$\nabla_{\mathbf{U}_0}\boldsymbol{\mathcal{J}} = \delta^{'}\mathbf{U}(t_0) \tag{2.20}$$

In summary, the gradient $\nabla_{\mathbf{U}_0}\boldsymbol{\mathcal{J}}$ can be obtained, for given $\mathbf{U}_0$ and $\mathbf{Y}$, by performing the following operations:

- Starting from $\mathbf{U}_0$ at time t_0 for state parameters $\mathbf{Y}$ we integrate the basic evolution equation (2.6) from t_0 to t_W; we store the values thus computed for $\mathbf{U}$ for $t_0 \leq t \leq t_W$.

- Starting from $\delta^{'}\mathbf{U}(t_W) = 0$ we integrate backwards in time (from t_W to t_0) the adjoint equation (2.17). The operator $\mathbf{F}^{'*}(t)$ and the gradient $\nabla_{\mathbf{U}}\mathbf{H}$ are determined, at each time t, from the values $\mathbf{U}(t)$ computed in the direct integration of (2.6).
- The final value $\delta^{'}\mathbf{U}(t_0)$ is the required gradient. $\nabla_{\mathbf{U}_0}\boldsymbol{\mathcal{J}}$

This adjoint approach was implemented in our research for two optimal control problems. The first problem has the objective of *suppressing the Karman vortices* for a flow around a rotating cylinder. The second problem, for the shock-tube problem, is related to the *change of discontinuity location* to a “desired” location.

We have studied the optimal control problems from a theoretical point of view (proving the existence of the solutions) as well as from a numerical perspective.

2.3 Sensitivity analysis

Sensitivity analysis (**SA**) studies the influence, quantitatively and qualitatively, of different internal or external parameters upon the model output (numerical and otherwise).

A general sensitivity theory for nonlinear systems was formulated by Cacuci [24], [25]. The physical problem under consideration is represented by the following system of K coupled nonlinear equations written in operator form as

$$\mathbf{N}[\mathbf{U}(\mathbf{X}), \mathbf{Y}(\mathbf{X})] = \mathbf{Q}[\mathbf{Y}(\mathbf{X}), \mathbf{X}] \tag{2.21}$$

where $\mathbf{X} = (X_1, \ldots, X_m) \subset \mathbb{R}^m$ is the position vector, $\mathbf{U}(\mathbf{X}) = [U_1(\mathbf{X}), \ldots, U_K(\mathbf{X})]$ is the state vector and $\mathbf{Y}(\mathbf{X}) = [Y_1(\mathbf{X}), \ldots, Y_P(\mathbf{X})]$ the vector of system parameters. $\mathbf{Q}[\mathbf{Y}(\mathbf{X}), \mathbf{X}]$ represents inhomogeneous source terms and the components of $\mathbf{N}[\mathbf{U}(\mathbf{X}), \mathbf{Y}(\mathbf{X})] = [N_1(\mathbf{U}, \mathbf{Y}), \ldots, N_K(\mathbf{U}, \mathbf{Y})]$ are nonlinear operators acting not only on the state vector $\mathbf{U}(\mathbf{X})$ but also on the vector of system parameters $\mathbf{Y}(\mathbf{X})$.

The system response $\mathbf{R} = \mathbf{R}(\mathbf{U}, \mathbf{Y})$ associated with the problem modeled by Eq. (2.21) must also be specified. The response considered here $\mathbf{R} = \mathbf{R}(\mathbf{e})$ is a general nonlinear functional of $\mathbf{e} = (\mathbf{U}, \mathbf{Y})$ with values in the set of real numbers.

The most general definition of a response to variations in the system parameters is the Gateaux differential (G-differential). The G-differential $\boldsymbol{\mathcal{V}}\mathbf{R}(\mathbf{e}^0; \mathbf{h})$ of $\mathbf{R}(\mathbf{e})$ at $\mathbf{e}^0$ with increment $\mathbf{h} = (h_{\mathbf{U}}, h_{\mathbf{Y}})$ is defined as

$$\boldsymbol{\mathcal{V}}\mathbf{R}(\mathbf{e}^0, \mathbf{h}) = \lim_{\epsilon \to 0} \frac{\mathbf{R}(\mathbf{e}^0 + \epsilon\mathbf{h}) - \mathbf{R}(\mathbf{e}^0)}{\epsilon} \tag{2.22}$$

A property of the G-differential is that $\mathbf{R}$ need not be continuous in $\mathbf{U}$ and/or $\mathbf{Y}$ for $\boldsymbol{\mathcal{V}}\mathbf{R}(\mathbf{e}^0; \mathbf{h})$ to exist at $\mathbf{e}^0 = (\mathbf{U}^0, \mathbf{Y}^0)$ (Cacuci [24]). This property will be employed for the sensitivities of a flow with discontinuities which are discussed in chapter 9.

Given the vector of changes $\mathbf{h}_{\mathbf{Y}}$ around $\mathbf{Y}^0$, the sensitivity $\boldsymbol{\mathcal{V}}\mathbf{R}(\mathbf{e}^0, \mathbf{h})$ at $\mathbf{e}^0$ can be evaluated only after determining the vector $\mathbf{h}_{\mathbf{U}}$, since $\mathbf{h}_{\mathbf{Y}}$ and $\mathbf{h}_{\mathbf{U}}$ are not independent. A relationship between $\mathbf{h}_{\mathbf{Y}}$ and $\mathbf{h}_{\mathbf{U}}$ is obtained by taking the G-differential of equation (2.21):

$$\boldsymbol{\mathcal{V}}\mathbf{N}(\mathbf{e}^0; \mathbf{h}) - \boldsymbol{\mathcal{V}}\mathbf{Q}(\mathbf{Y}^0; \mathbf{h}_{\mathbf{Y}}) = 0 \tag{2.23}$$

Once $\mathbf{h}_{\mathbf{U}}$ is determined it can be employed to evaluate the sensitivity $\boldsymbol{\mathcal{V}}\mathbf{R}(\mathbf{e}^0; \mathbf{h})$ of the response $\mathbf{R}(\mathbf{e})$ at $\mathbf{e}^0$.

We exemplify this approach for the same model discussed in the previous section

$$\frac{d\mathbf{U}(\mathbf{X})}{dt} = \mathbf{F}(t; \mathbf{U}(\mathbf{X}), \mathbf{Y}(\mathbf{X})) \tag{2.24}$$

Following Cacuci, sensitivity analysis can be applied to responses which are either functionals (i.e., scalar-valued operators) or operators (time-dependent or time/space dependent) of the model's parameters and variables. We present the case when the specific response is a functional of $\mathbf{U}$ and $\mathbf{Y}$.

In a similar way one approaches sensitivity analysis for responses which are operators (time-dependent or time/space dependent) of the model's parameters and variables.

We consider

$$\mathbf{R}(\mathbf{U}, \mathbf{Y}) = \int_{t_0}^{t_W} r(t; \mathbf{U}, \mathbf{Y})\, dt \tag{2.25}$$

where $r(t; \mathbf{U}, \mathbf{Y})$ depends on model variables $\mathbf{U}$, the parameters $\mathbf{Y}$ and the time interval $[t_0, t_W]$ represents the selected time window. The G-differential $\boldsymbol{\mathcal{V}}\mathbf{R}(\mathbf{U}^0, \mathbf{Y}^0; \mathbf{h_U}, \mathbf{h_Y})$ of the response function is given by

$$\boldsymbol{\mathcal{V}}\mathbf{R}(\mathbf{U}^0, \mathbf{Y}^0; \mathbf{h_U}, \mathbf{h_Y}) = \int_{t_0}^{t_W} r'_{\mathbf{U}} \cdot \mathbf{h_U}\, dt + \int_{t_0}^{t_W} r'_{\mathbf{Y}} \cdot \mathbf{h_Y}\, dt \tag{2.26}$$

where

$$r'_{\mathbf{U}} = \left(\frac{\partial r}{\partial U_1}, \dots, \frac{\partial r}{\partial U_K} \right)\Bigg|_{(\mathbf{U}^0, \mathbf{Y}^0)}$$
$$r'_{\mathbf{Y}} = \left(\frac{\partial r}{\partial Y_1}, \dots, \frac{\partial r}{\partial Y_P} \right)\Bigg|_{(\mathbf{U}^0, \mathbf{Y}^0)}$$

with K the dimension of the model parameters and P the dimension of the model variable $\mathbf{Y}$.

Taking the G-differential of (2.24) we obtain the linear system

$$\begin{aligned} \mathbf{L}(\mathbf{U}^0(t), \mathbf{Y}^0)\mathbf{h_U}(t) &= \mathbf{Q}(\mathbf{U}^0(t), \mathbf{Y}^0)\mathbf{h_Y}(t) \\ \mathbf{h_U}|_{t=t_0} &= 0 \end{aligned} \tag{2.27}$$

where

$$\mathbf{L}(\mathbf{U}^0(t), \mathbf{Y}^0) = \frac{d}{dt}\mathbf{I} - \frac{\partial \mathbf{F}}{\partial \mathbf{U}}$$
$$\mathbf{Q}(\mathbf{U}^0(t), \mathbf{Y}^0) = \frac{\partial \mathbf{F}}{\partial \mathbf{Y}}$$

and $\mathbf{I}$ is a unit matrix.

The value of $\mathbf{h_U}$ may be obtained by integrating (2.27) and, as discussed above, it can be employed to evaluate the sensitivity $\boldsymbol{\mathcal{V}}\mathbf{R}$. This approach is denoted as *forward sensitivity formalism.* However, when the dimension of the initial state vector and the number of parameters are large, the computational cost of calculating $\mathbf{h_U}$ is very high. Therefore we eliminate $\mathbf{h_U}$ by using the *adjoint sensitivity formalism.*

The adjoint operator $\mathbf{L}^*$ is defined through the relationship

$$\int_{t_0}^{t_W} \mathbf{h}_\mathbf{U} \cdot (\mathbf{L}^* \mathbf{q})\, dt = \int_{t_0}^{t_W} \mathbf{q} \cdot (\mathbf{L}\mathbf{h}_\mathbf{U})\, dt - [\mathbf{h}_\mathbf{U} \cdot \mathbf{q}]\Big|_{t_0}^{t_W} \tag{2.28}$$

where $\mathbf{q}$ is an arbitrary vector of dimension P.

Defining the adjoint model as

$$\begin{aligned} \mathbf{L}^* \mathbf{q} &= r'_\mathbf{U} \\ \mathbf{q}(t_W) &= 0 \end{aligned} \tag{2.29}$$

we write equation (2.28) as

$$\int_{t_0}^{t_W} r'_\mathbf{U} \cdot \mathbf{h}_\mathbf{U}\, dt = \int_{t_0}^{t_W} \mathbf{q} \cdot (\mathbf{L}\mathbf{h}_\mathbf{U})\, dt + \mathbf{h}_\mathbf{U}(t_0) \cdot \mathbf{q}(t_0) \tag{2.30}$$

Substituting (2.27) into (2.30) we get

$$\int_{t_0}^{t_W} r'_\mathbf{U} \cdot \mathbf{h}_\mathbf{U}\, dt = \int_{t_0}^{t_W} \mathbf{q} \cdot (\mathbf{Q}\mathbf{h}_\mathbf{Y})\, dt + \mathbf{h}_\mathbf{U}(t_0) \cdot \mathbf{q}(t_0) \tag{2.31}$$

A comparison (2.31) and (2.26) shows that

$$\boldsymbol{\mathcal{V}}\mathbf{R} = \int_{t_0}^{t_W} r'_\mathbf{Y} \cdot \mathbf{h}_\mathbf{Y}\, dt + \int_{t_0}^{t_W} \mathbf{q} \cdot (\mathbf{Q}\mathbf{h}_\mathbf{Y})\, dt + \mathbf{h}_\mathbf{U}(t_0) \cdot \mathbf{q}(t_0) \tag{2.32}$$

We note that the adjoint variable $\mathbf{q}(t)$ is the solution of the adjoint equations (2.29), which are independent of $\mathbf{h}_\mathbf{U}$ and $\mathbf{h}_\mathbf{Y}$.

The value of $\mathbf{h}_\mathbf{U}$, determined by the equation (2.27), does not depend on response and has to be computed only once. Therefore a single adjoint model calculation suffices to obtain the sensitivities to all the model parameters' variation. However, the forcing term $r'_\mathbf{U}$ in the adjoint model depends on the functional defining the response, so that for each response the adjoint equations model must be integrated again.

We conclude this section by mentioning that our research employed *local sensitivity analysis* as compared to *global sensitivity analysis* (Cacuci [26]) The objective of local sensitivity analysis is to analyze the behavior of the system responses locally around a chosen point or trajectory in the combined phase-space of parameters and state variables. On the other hand, the objective of global sensitivity analysis is to determine all of the system's critical points (namely bifurcations, turning points, extrema) in the combined phase-space formed by the parameters, state variables and adjoint variables and subsequently to analyze these critical points by local sensitivity analysis.

CHAPTER 3

NUMERICAL IMPLEMENTATION OF THE ADJOINT METHOD FOR COMPUTING THE GRADIENT OF THE COST FUNCTIONAL

In this chapter we present the numerical adjoint approach for the computation of the gradient of the cost functional with respect to the control parameters. First we describe a general form of the cost functional, which takes into account many additional influences: e.g., errors from observations, numerical computation errors or background terms.

3.1 The general expression of the cost functional

We recall that the collection of numbers needed to represent the state of the model is collected as a column matrix called the state vector $\mathbf{X}$. How the vector components relate to the real state depends on the choice of discretization, which is mathematically equivalent to a choice of basis.

One must distinguish between reality itself (which is more complex that what can be represented as a state vector) and the best possible representation of reality as a state vector, which we shall denote $\mathbf{X}_{true}$, the *true state* at the time of analysis. Another important value is $\mathbf{X}_{bg}$, the a priori or *background* estimate of the true state before the analysis is carried out, valid at the same time. Finally the analysis is denoted by $\mathbf{X}_{an}$ and this is what we are looking for.

In practice is often inconvenient to solve the analysis problem for all components of the model state. In these cases the work space of the analysis is not the model space but the space allowed for the corrections to the background. Then the analysis problem is to find a correction $\delta\mathbf{X}$ such that

$$\mathbf{X}_{an} = \mathbf{X}_{bg} + \delta\mathbf{X} \tag{3.1}$$

is as close as possible to $\mathbf{X}_{true}$.

For a given analysis we use a number of observed values. They are gathered into an observation vector $\mathbf{X}^{obs}$. To use them in the analysis procedure it is necessary to be able to compare them with the state vector. In practice it is very common that there are fewer observations than variables in the model and they are irregularly distributed, so that the only correct way to compare observations with the state vector is through the use of a function from model state space to observation space called an *observation operator* denoted by $\mathcal{H}$. This operator generates the values $\mathcal{H}(\mathbf{X})$ that the observations would take if both they and the state vector were perfect, in the absence of any modeling error. In practice $\mathcal{H}$ is a collection of interpolation operators from the model discretization to the observation points and conversions from model variables to the observed parameters. To evaluate the discrepancies between the observations and the state vector we consider the vector of departures at the observation points $\mathbf{X}^{obs} - \mathcal{H}(\mathbf{X})$.

In practice we may assume that there are errors between the above presented vectors and their true counterparts. They are modeled as follows:

- BACKGROUND ERRORS $\epsilon_{bg} = \mathbf{X}_{bg} - \mathbf{X}_{true}$. They are estimation errors of the background state, i.e. the difference between the background state vector vector and its true value. They do not include discretization errors.

- OBSERVATION ERRORS $\epsilon_{obs} = \mathbf{X}^{obs} - \mathcal{H}(\mathbf{X}_{true})$. They contain errors in the observation process (instrumental errors), errors in the design of the operator $\mathbf{H}$ and discretization errors which prevent $\mathbf{X}_{true}$ from being a perfect image of the true state

- ANALYSIS ERRORS $\epsilon_{an} = \mathbf{X}_{an} - \mathbf{X}_{true}$. They are estimation errors of the analysis state, which is what we want to minimize.

To represent the fact that there is some uncertainty in the background, in the observations and in the analysis we assume some *probability density function* for each kind of error. We can calculate statistics such as averages, variances and histograms of frequencies for the errors $\epsilon_{bg}, \epsilon_{obs}, \epsilon_{an}$.

As an example let us consider the case of background error ϵ_{bg}. If we were able to repeat each analysis experiment a large number of times, under exactly the same conditions, but with different realizations of errors generated by unknown cases, ϵ_{bg} would be different each time. In the limit of a very large number of realizations we expect the statistics to converge to values which depend only on the physical processes responsible for the errors and not on any particular realization of these errors.

For practical purposes some useful information on the average values of the statistics of errors can be gathered by different methods: for example, one can use forecast differences as surrogates to short-range forecast errors or one can estimate flow-dependent error covariances directly from a Kalman filter.

Uncertainty analysis methods can also be applied to ascertain the credibility of simulations using the numerical model. A review by Walters and Huyse [201] presents deterministic and probabilistic methods for uncertainty analysis.

We include the above mentioned terms in the expression of the cost functional. Over a given time interval, the analysis being at the initial time and the observations being distributed among nT times in the interval, we denote by the subscript i the quantities at any given observation time i. Hence $\mathbf{X}_i^{obs}, \mathbf{X}_i$ and $\mathbf{X}_{true,i}$ are the observations, the model and the true states at time i. $\mathbf{R}_i$ is the error covariance matrix for the observation errors $\mathbf{X}_i^{obs} - \mathcal{H}_i(\mathbf{X}_{true,i})$. The background error covariance matrix $\mathbf{B}$ is only defined at initial time (which is also the time of the background $\mathbf{X}_{bg}$ and of the analysis $\mathbf{X}_{an}$). We also consider a weighting matrix $\mathbf{W}$.

Then the cost functional will be written as a sum of a background term $\mathcal{J}_{BG}$ and an observation term $\mathcal{J}_{OBS}$:

$$\begin{aligned} \mathcal{J}(\mathbf{X}) &= \mathcal{J}_{BG}(\mathbf{X}) + \mathcal{J}_{OBS}(\mathbf{X}) \qquad (3.2) \\ &= (\mathbf{X} - \mathbf{X}_{bg})^T \mathbf{B}^{-1} (\mathbf{X} - \mathbf{X}_{bg}) + \sum_{i=0}^{nT} (\mathbf{X}_i^{obs} - \mathcal{H}_i(\mathbf{X}_i)^T \mathbf{W}_i \mathbf{R}_i^{-1} (\mathbf{X}_i^{obs} - \mathcal{H}_i(\mathbf{X}_i) \end{aligned}$$

Thus solving the optimal control problem is reduced to the minimization of the above cost functional subject to the strong constraint that the sequence of model states $\mathbf{X}_i$ must verify

$$\mathbf{X}_i = \mathcal{M}_{0\to i}(\mathbf{X}) \tag{3.3}$$

where $\mathcal{M}_{0\to i}$ is the numerical model from initial time to the i-th time.

Since our model is computed using time integration from initial time to the final time we may assume that the numerical model can be expressed as the product of intermediate operators $\mathcal{M}_i$

$$\mathbf{X}_{nT} = \mathcal{M}_{nT}\mathcal{M}_{nt-1}\cdots\mathcal{M}_1\mathbf{X}_0 \tag{3.4}$$

where nT corresponds to the final time and $\mathbf{X}_0$ is the initial condition.

We assume that we can linearize the operators $\mathbf{H}_i$ and $\mathcal{M}_{0\to i}$, i.e.,

$$\mathbf{X}_i^{obs} - \mathcal{H}_i\mathcal{M}_{0\to i}(\mathbf{X}) = \mathbf{X}_i^{obs} - \mathcal{H}_i\mathcal{M}_{0\to i}(\mathbf{X}_{bg}) - \mathbf{H}_i\mathbf{M}_{0\to i}(\mathbf{X} - \mathbf{X}_{bg}) \tag{3.5}$$

obtaining the *tangent linear model*. The existence of the tangent linear model depends on the model itself as well as on the length of the time interval considered.

3.2 Numerical gradient of the cost functional using the adjoint method

For simplicity, we employ a cost functional without the terms involved in error analysis to derive the algorithm of computing the numerical gradient of the cost functional using the adjoint method. Then we will present the a similar algorithm at the end of this section, this time for the general form of the cost functional (3.2).

Let us consider the cost functional as follows:

$$\boldsymbol{\mathcal{J}}[\mathbf{X},\Lambda] = \frac{1}{2}\sum_{k=0}^{nT}[\mathbf{X}(t_k) - \mathbf{X}^{obs}(t_k)]^T\mathbf{W}(t_k)[\mathbf{X}(t_k) - \mathbf{X}^{obs}(t_k)] \tag{3.6}$$

where $\mathbf{W}(t_k)$ a diagonal weighting matrix, Λ is the vector of control parameters, $t_0 \leq t_k \leq t_R$, $[t_0, t_{nT}]$ is the minimization window and nT is the number of time steps in the minimization window.

To find the minimum of the cost functional, efficient minimization algorithms require the calculation of the gradient of the cost functional with respect to the control parameters: $(\nabla_\Lambda\boldsymbol{\mathcal{J}}[\Lambda])^T$.

Near $\mathbf{X}(\tau)$ (the state vector at time τ) the nonlinear model can be written as:

$$\mathbf{X}(\tau + \Delta t) = \mathbf{F}(\mathbf{X}(\tau))].$$

To calculate the gradient of the cost functional with respect to the control parameters we define the change in the cost function resulting from a small perturbation $\delta\boldsymbol{\Lambda}$ about the model control parameters $\boldsymbol{\Lambda}$:

$$\delta\mathcal{J}[\mathbf{X},\Lambda] = \boldsymbol{\mathcal{J}}[\mathbf{X},\Lambda + \delta\boldsymbol{\Lambda}] - \boldsymbol{\mathcal{J}}[\mathbf{X},\Lambda] \tag{3.7}$$

As we take the limit $||\delta\mathbf{\Lambda}|| \to 0$, $\delta\mathbf{J}[X,\Lambda]$ is the directional derivative in the $\delta\mathbf{\Lambda}$ direction and it is given by:

$$\delta\boldsymbol{\mathcal{J}}[\mathbf{X},\Lambda] = \{\nabla_\Lambda \boldsymbol{\mathcal{J}}[\Lambda]\}^T \delta\mathbf{\Lambda} \tag{3.8}$$

On the other hand, $\delta\boldsymbol{\mathcal{J}}[\mathbf{X},\Lambda]$ may also be expressed in the following form (using definition (3.6) of the cost functional):

$$\delta\mathbf{J}[X,\Lambda] = \sum_{k=0}^{nT}(\mathbf{W}(t_k)[\mathbf{X}(t_k) - \mathbf{X}^{obs}(t_k)])^T \delta\mathbf{X}(t_k) \tag{3.9}$$

where $\delta\mathbf{X}(t_k)$ is the perturbation of the state vector obtained from the perturbation of the model parameters $\delta\Lambda$.

Combining relations (3.8) and (3.9) we obtain:

$$\{\nabla_\Lambda \boldsymbol{\mathcal{J}}[\mathbf{X},\Lambda]\}^T \delta\mathbf{\Lambda} = \sum_{k=0}^{nT}(\mathbf{W}(t_k)[\mathbf{X}(t_k) - \mathbf{X}^{obs}(t_k)])^T \delta\mathbf{X}(t_k) \tag{3.10}$$

From the above relation it is clear that we should express $\delta\mathbf{X}(t_k)$ as a function of $\delta\mathbf{\Lambda}$ in order to obtain an expression for $\nabla_\Lambda \boldsymbol{\mathcal{J}}[\mathbf{X},\Lambda]$.

We start by linearizing the model about the current model solution:

$$\delta\mathbf{X}(t_0 + \Delta t) = \frac{\partial \mathbf{F}(\mathbf{X})(t_0)}{\partial \Lambda}\delta\mathbf{\Lambda} \tag{3.11}$$

Using (3.11) for each time step we obtain:

$$\begin{aligned}
\delta\mathbf{X}(t_k) &= \mathbf{N}(t_k - \Delta t)\delta\mathbf{X}(t_k - \Delta t) \\
&= \mathbf{N}(t_k - \Delta t)\mathbf{N}(t_k - 2\Delta t)\delta\mathbf{X}(t_k - 2\Delta t) \\
&= \mathbf{N}(t_k - \Delta t)\mathbf{N}(t_k - 2\Delta t)\mathbf{N}(t_k - 3\Delta t)\delta\mathbf{X}(t_k - 3\Delta t) \\
&= \cdots \\
&= \mathbf{Q}_k \delta\Lambda
\end{aligned} \tag{3.12}$$

where $\mathbf{N}(t) \equiv \dfrac{\partial \mathbf{F}[X(t)]}{\partial \Lambda}$ and $\mathbf{Q}_k$ represents the result of applying all the operator matrices in the linear model to obtain $\delta\mathbf{X}(t_k)$ from $\delta\Lambda$.

With the relation $\delta\mathbf{X}(t_k) = \mathbf{Q}_k\delta\Lambda$, equation (3.10) becomes:

$$\nabla_\Lambda \boldsymbol{\mathcal{J}}[\mathbf{X},\Lambda] = \sum_{k=0}^{nT} \mathbf{Q}_k^T \mathbf{W}(t_k)[\mathbf{X}(t_k) - \mathbf{X}^{obs}(t_k)] \tag{3.13}$$

We define the adjoint equations for the adjoint variables $\hat{\mathbf{\Lambda}}^{(k)}$:

$$\hat{\mathbf{\Lambda}}^{(k)}(t_0) = \mathbf{Q}_k^T \hat{\mathbf{\Lambda}}^{(k)}(t_k), \text{ for } k = 1, \ldots, nT \tag{3.14}$$

If the adjoint variable $\hat{\mathbf{\Lambda}}^{(k)}(t)$ at time t_k is initialized as:

$$\hat{\mathbf{\Lambda}}^{(k)}(t_k) = \mathbf{W}(t_k)[\mathbf{X}(t_k) - \mathbf{X}^{obs}(t_k)]$$

then the gradient of the cost function with respect to the control parameters is:

$$\nabla_{\mathbf{\Lambda}}\boldsymbol{\mathcal{J}}[\mathbf{X}] = \sum_{k=0}^{nT} \hat{\mathbf{\Lambda}}^{(k)}(t_k)$$

Now we can write the algorithm for computing the gradient of the general cost functional (3.2).

The first stage is direct integration of the model from the initial time to the final time and the computation of the observation term in the cost functional:

1. Integration of the model from initial time to observation time i:

$$\mathbf{X}_i = \mathcal{M}_i\mathcal{M}_{i-1}\cdots\mathcal{M}_1\mathbf{X}_0 \tag{3.15}$$

2. Compute and store the "normalized departures"

$$d_i = \mathbf{R}_i^{-1}(\mathbf{X}^{obs} - \mathcal{H}_i(\mathbf{X}_i)) \tag{3.16}$$

3. Compute the contributions to the cost function

$$\boldsymbol{\mathcal{J}}_{OBS,i}(\mathbf{X}) = (\mathbf{X}^{obs} - \mathcal{H}_i(\mathbf{X}_i))^T d_i \tag{3.17}$$

4. Finally compute $\boldsymbol{\mathcal{J}}_{OBS}(\mathbf{X}) = \sum_{i=0}^{nT} \boldsymbol{\mathcal{J}}_{OBS,i}(\mathbf{X})$

The second stage is the computation of the gradient of the cost functional $\nabla\boldsymbol{\mathcal{J}}$. First we perform a slightly complex factorization of $\nabla\boldsymbol{\mathcal{J}}_{OBS}$:

$$\begin{aligned}
-\frac{1}{2}\nabla\boldsymbol{\mathcal{J}}_{OBS} &= -\frac{1}{2}\sum_{i=0}^{nT}\nabla\boldsymbol{\mathcal{J}}_{OBS,i} \\
&= \sum_{i=0}^{nT}\mathbf{M}_1^T\cdots\mathbf{M}_i^T\mathbf{H}_i d_i \\
&= \mathbf{H}_0^T d_0 + \mathbf{M}_1^T[\mathbf{H}_1^T d_1 + \mathbf{M}_2^T[\mathbf{H}_2^T d_2 + \cdots + \mathbf{H}_{nT}^T d_{nT}]\ldots]
\end{aligned}$$

where $\mathbf{M}_i$ and $\mathbf{H}_i$ are the linearized operators corresponding to $\mathcal{M}_i$ and $\mathcal{H}_i$.

The last expression is evaluated from right to left in the following steps:

1. Initialize the *adjoint* variable $\hat{\mathbf{X}}$ to zero at final time: $\hat{\mathbf{X}}_{nT} = 0$.

2. For each time step $i-1$ the variable $\hat{\mathbf{X}}_{i-1}$ is obtained by adding the *adjoint forcing* $\mathbf{H}_i^T d_i$ to $\hat{\mathbf{X}}_i$ and by performing the *adjoint integration* by multiplying the result by $\mathbf{M}_i^T$, i.e.

$$\hat{\mathbf{X}}_{i-1} = \mathbf{M}_i^T(\hat{\mathbf{X}}_i + \mathbf{H}_i^T d_i) \tag{3.18}$$

3. At the end of the recurrence the value of the adjoint variable $\hat{\mathbf{X}}_0$ gives the required result for the gradient of the observation term in the cost functional

$$\hat{\mathbf{X}}_0 = -\frac{1}{2}\nabla\boldsymbol{\mathcal{J}}_{OBS}(\mathbf{X}) \tag{3.19}$$

Finally add the gradient of the background term to compute the numerical value of the gradient of the cost functional (3.2)

$$\nabla\boldsymbol{\mathcal{J}}(\mathbf{X}) = 2\mathbf{B}^{-1}(\mathbf{X} - \mathbf{X}_{bg}) + \nabla\boldsymbol{\mathcal{J}}_{OBS}(\mathbf{X}) \tag{3.20}$$

The terminology employed in the algorithm reflects the fact that the computations look like the integration of an *adjoint model* backward in time with a time-stepping defined by the transpose time-stepping operators $\mathbf{M}_i^T$ and an external forcing $\mathbf{H}_i^T d_i$, which depends on the distance between the model trajectory and the observations.

3.3 Coding the adjoint and the tangent linear method

If we linearize the nonlinear model we obtain the tangent linear model (**TLM**). The transpose of the **TLM** is the adjoint model.

For coding the **TLM**, we linearize the original nonlinear forward model code line by line, **DO** loop by **DO** loop and subroutine by subroutine.

If we view the tangent linear model as the result of the multiplication of a number of operator matrices: $\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_M$ where each matrix $A_i, (i = 1, \ldots, M)$ represents either a subroutine or a single DO-loop, then the adjoint model can be viewed as being a product of adjoint subproblems: $\mathbf{A}_M^T\mathbf{A}_{M-1}^T\cdots\mathbf{A}_1^T$.

The correctness of the adjoint of each operator was checked using the following identity:

$$(\mathbf{AQ})^T(\mathbf{AQ}) = \mathbf{Q}^T(\mathbf{A}^T(\mathbf{AQ}))$$

where **Q** represents the input of the original code and **A** can be either a single DO loop or a subroutine. All subroutines of the adjoint model were subjected to this test.

3.4 Accuracy of the gradient for the continuous cost functional

The accuracy of the gradients calculated by the adjoint method should be at the level of machine precision. Errors could result due to coding mistakes, round-off errors or the presence of non differentiable functions.

A method for the gradient check is described below, using the following Taylor expansion of the cost functional:

$$\boldsymbol{\mathcal{J}}(\mathbf{X} + \eta\mathbf{h}) = \boldsymbol{\mathcal{J}}(\mathbf{X}) + \eta\mathbf{h}^T\nabla\boldsymbol{\mathcal{J}}(\mathbf{X}) + \mathcal{O}(\eta^2) \tag{3.21}$$

where $||\mathbf{h}|| = 1$, η scalar and $\nabla\boldsymbol{\mathcal{J}}(\mathbf{X})$ is the gradient of the cost functional $\boldsymbol{\mathcal{J}}(\mathbf{X})$ with respect to **X** computed using the adjoint code.

Rewriting the above formula, a function of η can be defined as (see Navon et al. [148]):

$$\mathbf{\Phi}(\eta) = \frac{\boldsymbol{\mathcal{J}}(\mathbf{X} + \eta\mathbf{h})) - \boldsymbol{\mathcal{J}}(\mathbf{X})}{\eta\mathbf{h}^T\nabla\boldsymbol{\mathcal{J}}(\mathbf{X})} \tag{3.22}$$

The gradient computed using the adjoint model can be assumed to be completely accurate (up to the machine error) when $\lim_{\eta\to 0} |\mathbf{\Phi}(\eta)| = 1$. A validity region of the gradient test is normally obtained for $10^{-3} \geq \eta \geq \epsilon$ (where ϵ is the machine accuracy). For $\eta > 10^{-3}$ we have truncation error and for η near the machine accuracy roundoff errors prevail.

The results of the gradient check test are displayed in Fig. 3.1.

3.5 Accuracy of the gradient of the non smooth cost functional

If the cost functional is non smooth the *gradient* of the cost functional does not exist everywhere. In this case the adjoint method computes a *subgradient* of the cost functional.

We considered the subgradient obtained from the adjoint model to be sufficiently accurate if the following tolerances were satisfied

$$\lim_{\eta\to 0} |\mathbf{\Phi}(\eta)| = \delta \tag{3.23}$$

for $10^{-3} \geq \eta \geq 10^{-10}$
where δ is a constant number which depends on the problem parameters. Fig. 3.1 presents the subgradient check test.

We can see that the subgradient ratio tends to a constant number, a number which decreases slightly as we increase the time window.

3.6 Checkpointing

According to theoretical bounds (e.g., Griewank [80]) the reverse mode of differentiation allows the generation of an adjoint code involving at most five times the number of operations of the original model. However this low operation count requires the storage of the full trajectory, which is formed by the values of the variables of the original model that may be used for the evaluation of linearized statements. The calculation of the trajectory has to be performed before (or during) the backward integration of the adjoint code. When the differentiated codes require too large an amount of memory, a possible solution to alleviate this problem is to implement a *checkpointing* algorithm (Griewank [81], [84]), which provides an optimal logarithmic behavior in terms of time and memory requirements.

The strategy to solve the trajectory problem arising in adjoint computations is based on “divide et impera”. Griewank [81] proposed to save the state of the system from time to time during runs of the original code.

These are called *checkpoints* and they allow for the computation of parts of the trajectory without systematically coming back to the initial point. Checkpoints are stored on a stack in a *Last-In-First-Out* manner. Forward sweeps and reverse sweeps are then done, part by part, from checkpoints.

An example of a *checkpointing* algorithm is presented below:

1. **run** the original code and **store** the checkpoints
2. **for all** the checkpoints, taken in reverse order
    (a) **restore** the checkpoint
    (b) perform a **forward sweep** from the checkpoint to the previously removed one (or to a given iteration)
    (c) perform a **reverse sweep** down to the checkpoint
    (d) **remove** the checkpoint from the stack

Checkpointing is easy to implement for time-stepping problems, where a natural point is the beginning of a timestep. Using *checkpointing* introduces extra forward steps of the original model.

A practical implementation faces two challenges. The first task is the selection of checkpoints. The question is at which points of the whole computational process one should place the checkpoints to achieve an optimal reduction of the storage requirement. The second task is to manage all the information at every checkpoint that is necessary. This requires to save the system state at the checkpoint and to restore it in order to repeat the next computational steps, when that becomes necessary. Also the values of the adjoints must be managed to perform the successive reverse sweeps. This is a trade-off between memory and CPU time.

From the user's point of view the choice of a checkpointing scheme depends essentially on the particular code and the particular computer architecture the user deals with. A large variety of checkpointing schedules are discussed in the research of Charpentier [27] and Restrepo et al. [164].

3.7 Automatic differentiation

Automatic differentiation **AD** is a technique for augmenting computer programs with derivative computations. It exploits the fact that every computer program, no matter how complicated, executes a sequence of elementary arithmetic operations such as additions or elementary functions such as exp(). By applying the chain rule of derivative calculus repeatedly to these operations, derivatives of arbitrary order can be computed automatically and accurate to working precision.

In contrast to other methods like finite-difference gradients **AD** computes the exact derivative of the given code without any additional truncation error and without much additional theoretical work. Any computer code can be viewed as a concatenation of many evaluations of the intrinsic operators and functions. After finding the computational graph from the specified independent variable to the dependent variables, the derivative of such a concatenation can be obtained by applying the chain rule. This means formally the multiplications of all Jacobians of the intrinsic functions.

The main techniques in **AD** are described in Griewank and Corliss [83], Berz et al. [16] and Griewank [82]. The given source code is either transformed into a new code computing the desired derivatives (by tools working with **source transformation**) or linked with libraries that include overloaded versions of the intrinsic functions and operators of th used programming language (by tools based on **operator overloading**).

Based on how the chain rule is used to propagate derivatives through the computation two approaches to automatic differentiation have been developed: *the forward mode* for which the chain rule is applied from the beginning to the end of the "active" section of the program and, respectively, *the reverse mode* if the computation is going from the end back to the beginning.

The forward mode propagates derivatives of intermediate variables with respect to the independent variables. Let us assume that $\mathbf{X} = (X_1, \ldots, X_n)$ are the independent variables and $\mathbf{Y} = (Y_1, \ldots, Y_m)$ are the dependent variables. The linearity of differentiation allows the forward mode to compute arbitrary linear combinations $\mathbf{J} * \mathbf{S}$, where $\mathbf{S}$ is a $n \times p$ matrix and $\mathbf{J}$ is the Jacobian

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial Y_1}{X_1} & \ldots & \dfrac{\partial Y_1}{X_n} \\ \ldots & \ldots & \ldots \\ \dfrac{\partial Y_m}{X_1} & \ldots & \dfrac{\partial Y_m}{X_n} \end{pmatrix} \tag{3.24}$$

The effort required is roughly $\mathcal{O}(p)$ times the runtime and memory of the original program. In particular, when $\mathbf{S}$ is a vector $\mathbf{s}$, we compute the directional derivative

$$\lim_{h \to 0} \frac{F(\mathbf{X} + h * \mathbf{s}) - F(\mathbf{X})}{h} \tag{3.25}$$

This type of differentiation is also used to obtain the tangent linear model.

The reverse mode of automatic differentiation propagates derivatives of the final result with respect to an intermediate quantity called the adjoint quantity. To propagate the adjoint one must be able to reverse the flow of the program and must remember or recompute any intermediate values that nonlinearly impacts the final result.

For a matrix $q \times m$ $\mathbf{W}$, the reverse mode allows us to compute the row linear combination $\mathbf{W} * \mathbf{J}$ with $\mathcal{O}(q)$ times as many floating-point operations as required for the evaluations of F. The storage requirements are harder to predict and depend to a large extent on the nonlinearity of the program and the implementation approach chosen. The reverse mode also corresponds to a method for obtaining the discrete adjoint.

In general the forward mode is appropriate if the number of independent variables is higher than the number of the dependent variables and the reverse mode in the contrary case.

The most important scientific **AD** codes are the following:

1. $\mathcal{ADIFOR/ADIC}$ available from Argonne National Laboratory/Rice University for Fortran77/C

2. $\mathcal{ODYSSEE}$ available from INRIA for Fortran77

3. $\mathcal{TAMC}$ by Ralf Giering for Fortran77/Fortran90

4. $\mathcal{ADOL} - C$ available from TU Dresden for C++

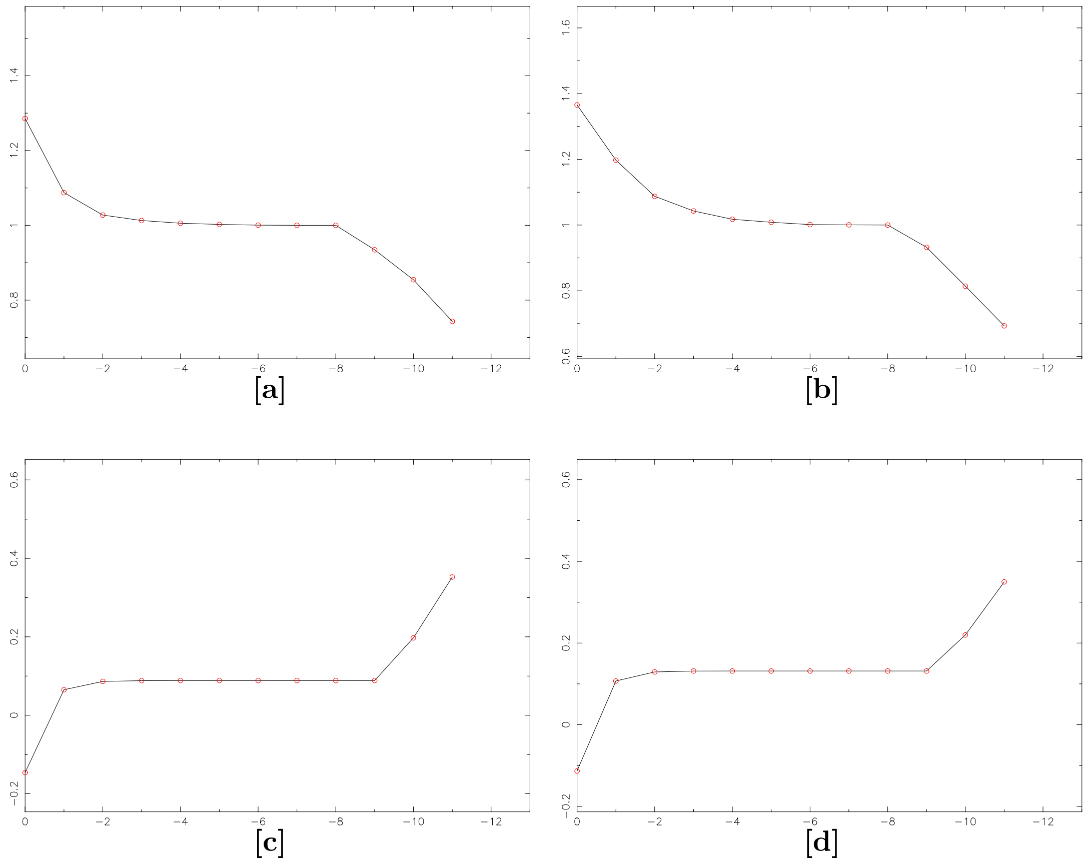


Figure 3.1. The accuracy check: the gradient of the cost functional vs. $\log(\eta)$ for the flow around the cylinder in the constant rotation case ([**a**]) and, respectively, time-dependent rotation case ([**b**]); a subgradient of the cost functional vs. $\log(\eta)$ for the shock-tube flow at time=0.24 for **AVM** model ([**c**]) and, respectively, **HRM** model ([**d**])

CHAPTER 4

OPTIMIZATION ALGORITHMS

Consider the following nonlinear constrained optimization problem

$$\begin{aligned} &\text{minimize } F(X) \\ &\text{subject to } X \in G \end{aligned} \tag{4.1}$$

where the objective function $F : \mathbf{R}^n \to \mathbf{R}$ is a locally LIPSCHITZ function on the feasible set $G \subset \mathbf{R}^n$ (if $G = \mathbf{R}^n$ then the problem is unconstrained).

A general iterative algorithm to solve the problem (4.1) is as follows

- Step 0. INITIALIZATION. Find a feasible starting point $X_1 \in G$ and set $k = 1$
- Step 1. DIRECTION FINDING. Find a feasible descent direction $d_k \in \mathbf{R}^n$

$$F(X_k + td_k) < F(X_k) \quad \text{and } X_k + td_k \in G \quad \text{for some } t > 0 \tag{4.2}$$

- Step 2. STOPPING CRITERION. If X_k is "close enough" to the required solution then STOP
- Step 3. LINE SEARCH. Find a step size $t_k > 0$ such that

$$t_k = \arg\min_{t>0} F(X_k + td_k) \text{ and } X_k + t_k d_k \in G \tag{4.3}$$

- Step 4. UPDATING. Set $X_{k+1} = X_k + t_k d_k$ and go to Step 1

4.1 Non differentiable minimization

If the function F to be minimized is non smooth then methods of non differentiable optimization are required. They can be divided into two main classes: subgradient methods and bundle methods.

Since the gradient of a non smooth function F exists only almost anywhere we have to replace the gradient by the generalized gradient

$$\begin{aligned} \partial F(X) = \text{conv}\{g| \text{ there exists a sequence } (X_i)_{i\in\mathbf{N}} \text{ such that } \lim_{i\to\infty} X_i = X, \\ F \text{ differentiable at } X_i, i \in \mathbf{N}, \text{ and } \lim_{i\to\infty} \nabla F(X_i) = g\} \end{aligned}$$

where "conv" stands for convex hull and it is defined as the closure of the set which contains all convex linear combinations of subgradients (an element of the generalized gradient is called subgradient).

The non smooth optimization methods are based on the assumptions that the function F is locally Lipschitz continuous and we can evaluate the function and its arbitrary subgradient at each point.

4.1.1 The subgradient methods

The history of subgradient methods starts in the 60s: Shor (1962), Polyak (1964), Ermolev (1967).

The main idea is to employ only one subgradient $\xi_k \in \partial F(X_k)$ instead of the gradient $\nabla F(X_k)$. Hence the natural generalization of gradient method is to replace the gradient by the normalized gradient in the formula for d_k defined in Step 2:

$$d_k = -\xi_k/||\xi_k|| \tag{4.4}$$

The above strategy of generating d_k do not ensure descent and hence minimizing line searches becomes unrealistic. Also the standard stopping criterion can no longer be applied since an arbitrary subgradient contains no information on the optimality condition $0 \in \partial F(X)$.

Due to these facts we are forced to use a priori choice of step sizes t_k to avoid line searches and the stopping criterion. Thus we define the next iteration point by

$$X_{k+1} = X_k - t_k \frac{\xi_k}{||\xi_k||} \tag{4.5}$$

where $\xi_k \in \partial F(X_k)$ and a suitable $t_k > 0$ was chosen.

In order to accelerate the rate of convergence we may try to generalize more smooth methods than the gradient method. The most efficient methods at the moment are based on generalized Quasi-Newton methods: ellipsoid and space dilation algorithms by Shor [175] and the variable metric method by Uryasev [196].

4.1.2 The bundle methods

The guiding principle behind them is to exploit the previous iterations by gathering the subgradient information into a *bundle* of subgradients. The pioneering bundle method, the ϵ-steepest descent method, was developed by Lemarechal [125]. The main difficulty in Lemarechal's method is the a priori choice of an approximation tolerance which controls the radius of the ball in which the bundle model is thought to be a good approximation of the objective function.

A different approach was presented by Kiwiel [121], based on the cutting plane method. The basic idea is to form a convex piecewise linear approximation to the objective function using the linearizations generated by subgradients. Kiwiel also presented two strategies to bound the number of stored subgradients: subgradient selection and aggregation. The main disadvantage of Kiwiel's method is its sensitivity to scaling of the objective function. Also the uncertain line search may require, in general, many function evaluations compared with the number of iterations.

In spite of different backgrounds, both methods (Lemarechal and Kiwiel) generate the search direction at each direction by solving quadratic detection finding problems.

A new approach that combines the bundle idea with the trust region method was adopted by Schramm and Zowe (Bundle Trust Region method [171]) and by Kiwiel (Proximal Bundle method [122]).

The following two features are characteristic to bundle methods:

- the gathering of subgradient information from past iterations into a bundle
- the concept of SERIOUS STEP and NULL STEP in line search

Let $Y_{k+1} = X_k + t_k d_k$ for some $t_k > 0$ and $\xi_{k+1} \in \partial F(Y_{k+1})$. Then we have the following sequence:

1. Make a SERIOUS STEP $X_{k+1} = Y_{k+1}$ if
   $F(Y_{k+1}) \leq F(X_k) - \delta_k$ for some $\delta_k > 0$; add ξ_{k+1} into bundle
2. Otherwise make a NULL STEP $X_{k+1} = X_k$; add ξ_{k+1} into bundle

4.1.3 The hybrid algorithm (PVAR) for nonsmooth minimization

The most efficient globally convergent algorithms for nonconvex non smooth optimization are based on versions of the bundle methods (e.g. Lemarechal [126], Bonnans et al. [20], Schramm and Zowe [171], Makela and Neittaanmaki [138]). We employed a hybrid method (described in Vlcek and Luksan [198] and Luksan and Vlcek [137]) which combines the characteristics of the variable metric method and the bundle method.

The algorithm generates a sequence of basic points $(x_k)_{k\in\mathbf{N}}$ and a sequence of trial points $(y_k)_{k\in\mathbf{N}}$ satisfying

$$x_{k+1} = x_k + t_L^k d_k, \qquad y_{k+1} = x_k + t_R^k d_k$$

with $y_1 = x_1$, where $t_R^k \in (0, t_{\max}], t_L^k \in [0, t_R^k]$ are appropriately chosen step sizes, $d_k = -H_k \tilde{g_k}$ is a direction vector and $\tilde{g_k}$ is an aggregate subgradient.

The matrix H_k accumulates information about the previous subgradients and represents an approximation of the inverse Hessian matrix if the function F is smooth.

If the descent condition $F(y_{k+1}) \leq F(x_k) - c_L t_R^k w_k$ is satisfied with suitable t_R^k, where $c_L \in (0, 0.5)$ is fixed and $-w_k < 0$ represents the desirable amount of descent, then $x_{k+1} = y_{k+1}$ (descent step).

Otherwise a null step is taken which keeps the basic points unchanged but accumulates information about the minimized function.

The construction of the aggregate subgradient is presented below.

Let us denote by m the lowest index j satisfying $x_j = x_k$ (index of the iteration after last descent step).

We define $\tilde{g}_{k+1}$ as a convex combination of the following (known) subgradients: the basic subgradient $g_m \in \partial f(x_k)$, the trial subgradient $g_{k+1} \in \partial f(y_{k+1})$, and the current aggregate subgradient $\tilde{g}_k$

$$\tilde{g}_{k+1} = \lambda_{k,1} g_m + \lambda_{k,2} g_{k+1} + \lambda_{k,3} \tilde{g}_k$$

The multipliers λ_k can be determined easily by minimizing a simple quadratic function which depends on these three subgradients and two subgradient locality measures (this approach replaces the solution of a rather complicated quadratic programming problem which appears in the standard bundle method Lemarechal [126]).

The matrices H_k are generated using a symmetric quasi-Newton rank-one update after the null steps (to preserve the property of being bounded and other characteristics required for the global convergence) or the standard BFGS update after the descent steps.

For a more in-depth discussion about both types of updates the reader is referred to Fletcher [57].

4.2 Differentiable optimization

4.2.1 The Q-N algorithm for unconstrained minimization

A Quasi-Newton (**Q-N**) algorithm (also called variable metric method) was employed for the minimization of the cost functional for the flow past a cylinder. Instead of obtaining an estimate of the Hessian matrix at a single point, this method gradually builds up an approximate Hessian matrix by using gradient information from some or all of the previous iterates visited by the algorithm.

We started with the identity matrix and then a better approximation $\mathbf{H}_k$ to the inverse Hessian matrix was built up, iteratively, in such a way that the matrix $\mathbf{H}_k$ preserves positive definiteness and symmetry.

Given the current iterate x_i, and the approximate Hessian matrix H_k at x_k , the linear system

$$H_k\mathbf{x}_k = -\nabla\boldsymbol{\mathcal{J}}(\mathbf{x}_k) \tag{4.6}$$

is solved to generate a direction p_k. The next iterate is then found by performing a line search along p_k and setting

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{H}_{k+1}\cdot(\nabla\boldsymbol{\mathcal{J}}(\mathbf{x}_{k+1}) - \nabla\boldsymbol{\mathcal{J}}(\mathbf{x}_k)) \tag{4.7}$$

where the new approximation to the inverse Hessian $\mathbf{H}_{k+1}$ is constructed using using the *Davidon-Fletcher-Powell* (DFP) rank-2 update formula.

We can make H_{k+1} to mimic the behavior of $\nabla^2\boldsymbol{\mathcal{J}}$ by enforcing the Quasi-Newton condition

$$H_{i+1}\mathbf{s}_i = \mathbf{y}_i \tag{4.8}$$

where $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ and $\mathbf{y}_k = \nabla\boldsymbol{\mathcal{J}}(\mathbf{x}_{k+1}) - \nabla\boldsymbol{\mathcal{J}}(\mathbf{x}_k)$.

This condition can be satisfied by making a simple low-rank update to H_k. The most commonly used family of updates is the Broyden class of rank-two updates, which have the form

$$H_{k+1} = H_k - \frac{H_k\mathbf{s}_k(H_k\mathbf{s}_k)^T}{\mathbf{s}_k^T H_k\mathbf{s}_k} + \frac{\mathbf{y}_k\mathbf{y}_k^T}{\mathbf{y}_k^T\mathbf{s}_k} + \Phi_k[\mathbf{s}_k^T H_k\mathbf{s}_k]\mathbf{v}_k\mathbf{v}_k^T \tag{4.9}$$

where $\Phi_k \in [0,1]$ and

$$\mathbf{v}_k = \frac{\mathbf{y}_k}{\mathbf{y}_k^T\mathbf{s}_k} - \frac{H_k\mathbf{s}_k}{\mathbf{s}_k^T H_k\mathbf{s}_k} \tag{4.10}$$

The choice $\Phi_k = 0$ gives the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update. The Davidon-Fletcher-Powell update, which was proposed earlier, is obtained by setting $\Phi_k = 1$.

We employed a modified version of the backtracking strategy implemented in Numerical Recipes [162] to choose a step along the direction of the Newton step $\mathbf{p}$. The goal was to move to a new point x_{new} along the direction of the Newton step $\mathbf{p}$:

$$\mathbf{x}_{new} = \mathbf{x}_{old} + \lambda \mathbf{p}, \qquad 0 < \lambda \leq \lambda_0 \leq 1$$

such that the function

$$g(\lambda) = (\boldsymbol{\mathcal{J}} \mathbf{x}_{old} + \lambda \mathbf{p})$$

showed a sufficient decrease.

The convergence criteria used here are

$$\boldsymbol{\mathcal{J}}(\mathbf{x}_{new}) \leq \boldsymbol{\mathcal{J}}(\mathbf{x}_{old}) + \sigma \nabla \boldsymbol{\mathcal{J}} \cdot (\mathbf{x}_{new} - \mathbf{x}_{old}), \qquad 0 < \sigma < 1$$

or $||\nabla \boldsymbol{\mathcal{J}}(\mathbf{x}_{new})|| < 10^{-5}$.

4.2.2 The L-BFGS unconstrained optimization algorithm

We also implemented the **L-BFGS** method (Nocedal [151], Liu and Nocedal [134], Nocedal and Wright [152]) which performs the unconstrained minimization of a smooth nonlinear function for which the gradient is available. **L-BFGS** is a limited memory method based on the well-known BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm.

The main idea of this method is to use curvature information only from the most recent iterations to construct the Hessian approximation. Instead of storing fully dense $n \times n$ approximations, this approach saves just a few vectors (of length n) that represent the approximations implicitly.

Each step of the original **BFGS** method has the form

$$x_{k+1} = x_k - \alpha_k H_k \nabla \boldsymbol{\mathcal{J}}_k, \quad k = 0, 1, 2, \ldots$$

where α_k is the step length and $\boldsymbol{\mathcal{J}}$ is the cost functional. H_k is updated at each iteration by means of the formula

$$H_{k+1} = V_k^T H_k V_k + \beta_k s_k s_k^T \tag{4.11}$$

where

$$\beta_k = \frac{1}{y_k^T s_k} \qquad V_K = I - \beta_k y_k s_k^T \tag{4.12}$$

and

$$s_k = x_{k+1} - x_k \qquad y_k = \nabla \boldsymbol{\mathcal{J}}_{k+1} - \nabla \boldsymbol{\mathcal{J}}_k \tag{4.13}$$

$\boldsymbol{\mathcal{J}}_k$ being the cost functional at step k of the minimization iteration.

We say that the matrix H_{k+1} is obtained by updating H_k using the pair (s_k, y_k). For **L-BFGS** a modified version of H_k is stored implicitly, by using a certain number (say m) of the vector pairs (s_l, y_l) that are used in the formulae (4.11)-(4.13).

The product $H_k \boldsymbol{\mathcal{J}} \nabla_k$ can be obtained by performing a sequence of inner products and vector summations involving $\nabla \boldsymbol{\mathcal{J}}_k$ and the pairs (s_l, y_l). After the new iterate is computed, the oldest vector pair in the set of pairs (s_l, y_l) is deleted and replaced by the new pair (s_k, y_k) obtained from the current step (4.13). In this way the set of vector pairs includes curvature information from the m most recent iterations (usually $3 \leq m \leq 10$).

For numerical experiments using the **L-BFGS** method the reader is referred to Zou et al. [213].

We would like to conclude this section discussing our preference for **L-BFGS** over other smooth minimization algorithms. One may argue that for our case the number of control parameters may not justify the selection of a limited memory method.

While this may be true, we consider that our approach (using the adjoint method for the gradient computation) may be easily and successfully implemented for optimal control problems with a much greater number of control variables. In that case improvements in the efficiency of the numerical optimization will be determined not only by choosing the adjoint method over other methods for the gradient calculation but also by selecting a limited memory minimization algorithm.

4.2.3 Sequential Quadratic Programming SQP for constrained optimization

One of the most effective methods for nonlinearly constrained optimization is to generate steps by solving quadratic problems. This sequential quadratic programming (**SQP**) approach can be used both in line-search and trust region frameworks and it is appropriate for small or large problems.

Although we did not employ it in our research, it is described since it serves as an efficient minimization algorithm in large optimal control applications. A version of **SQP** coupled with trust-region methods and interior-point techniques was implemented in the package **TRICE** [45].

Let us consider an equality-constrained problem

$$\min F(X) \tag{4.14}$$

$$\text{subject to } C(X) = 0 \tag{4.15}$$

where $F : \mathbf{R}^n \to \mathbf{R}$ and $C : \mathbf{R}^n \to \mathbf{R}^m$ are smooth functions.

The essential idea of **SQP** is to model (4.14)-(4.15) at the current iterate X_k by a quadratic programming subproblems and to use the minimizer of this subproblem to define a new iterate X_{k+1}.

The challenge is how to design the quadratic subproblem so that it yields a good step for the underlying constrained optimization problem while the overall **SQP** algorithm has good convergence properties and good practical performance.

We denote by $\mathcal{L}(X, \lambda) = F(X) - \lambda^T C(X)$ the Lagrangian.

A is the Jacobian matrix of the constraints

$$\mathbf{A}(X)^T = [\nabla C_1(X), \nabla C_2(X), \ldots, \nabla C_m(X)] \tag{4.16}$$

and by $\mathbf{W}(X, \lambda) = \nabla^2_{XX} \mathcal{L}(X, \lambda)$ the Hessian of the Lagrangian.

At iteration (X_k, λ_k) we define the quadratic problem

$$\min_p \frac{1}{2} p^T W_k p + \nabla F_k^T p \tag{4.17}$$

$$\text{subject to } A_k p + C_k = 0$$

with A_k and W_k the approximations for **A** and respectively **W**.

If the constraint Jacobian A_k has full row rank and the matrix W_k satisfies $d^T W_k d > 0$ on the tangent space of constraints (i.e. for all $d \neq \mathbf{0}$ such that $A_k d = 0$) then the problem (4.17) has a unique solution (p_k, μ_k) that satisfies

$$\begin{aligned} W_k p + \nabla F_k^T p - A_k^T \mu_k &= 0 \\ A_k p_k + C_k &= 0 \end{aligned}$$

If p is the vector of descent and λ_{k+1} is the step for descent obtained from solving the system

$$\begin{bmatrix} W_k & -A_k^T \\ A_k & \mathbf{0} \end{bmatrix} \begin{bmatrix} p \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} -\nabla F_k \\ C_k \end{bmatrix} \tag{4.18}$$

then it can be shown that $p = p_k$ and $\lambda = \mu_k$.

To be practical, an **SQP** method must be able to converge from remote starting points and on nonconvex problems. If W_k is positive definite on the tangent space of constraints, the quadratic subproblem (4.17) can be solved without any additional considerations. When W_k does not have this property, line-search methods either replace it by a positive definite approximation B_k or modify W_k directly during the process of matrix factorization. Another approach is given by the trust-region methods, which add a constraint to the subproblem, limiting the step to a region where the model (4.17) is considered to be reliable.

Complications may arise, however, because the inclusion of the trust region may cause the subproblem to become infeasible. At some iterations it is necessary to relax the constraints, which complicates the algorithm and increases its computational cost. Due to these trade-offs, neither one of the two **SQP** approaches (line-search or trust region) can be regarded as clearly superior to the other.

Let us now consider the choice of the matrix W_k in the quadratic model. Various implementations of **SQP** based on specific choices of W_k have performed well on many problems. They yielded poor performance or even failure for other problems, however.

For this reason there is not a unique choice for W_k. We present here some of the most employed choices of W_k, based on Nocedal and Wright [152].

The first choice is based on maintaining a quasi-Newton approximation $\mathbf{B}_k$ to the full Hessian of the Lagrangian $\nabla^2_{XX}\mathcal{L}(X_k, \lambda_k)$ using a BFGS update. The update for $\mathbf{B}_k$ makes use of vectors s_k and Y_k

$$s_k = X_{k+1} - X_K \qquad\qquad Y_k = \nabla_X \mathcal{L}(X_{k+1}, \lambda_{k+1}) - \nabla_X \mathcal{L}(X_k, \lambda_{k+1}) \tag{4.19}$$

The new approximation $\mathbf{B}_{k+1}$ is then computed using the BFGS formula. For this approach the iteration will converge robustly and rapidly. If, however, $\nabla^2_{XX}\mathcal{L}$ contains negative eigenvalues the BFGS approach of approximating it with a positive matrix may be ineffective.

A more effective modification is the damped BFGS updating which ensures that the update is always well-defined by modifying the definition of Y_k. If we define s_k and Y_k as in (4.19) and set

$$r_k = \theta_k Y_k + (1 - \theta_k)\mathbf{B}_k s_k \tag{4.20}$$

where the scalar θ_k is defined as

$$\theta_k = \begin{cases} 1 & \text{if } \; s_k^T Y_k \geq 0.2 s_k^T \mathbf{B}_k s_k \\ (0.8 s_k^T \mathbf{B}_k s_k)/(s_k^T \mathbf{B}_k s_k - s_k^T Y_k) & \text{if } \; s_k^T Y_k < 0.2 s_k^T \mathbf{B}_k s_k \end{cases}$$

$\mathbf{B}_k$ is updated as follows

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k \mathbf{s}_k (\mathbf{B}_k \mathbf{s}_k)^T}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k} + \frac{r_k r_k^T}{s_k^T r_k} \tag{4.21}$$

which guarantees that $\mathbf{B}_{k+1}$ is positive definite.

But this method still fails to address the underlying problem that the Lagrangian Hessian may not be positive definite.

A different approach modifies the Lagrangian Hessian directly by adding terms to the Lagrangian function, the effect of which is to ensure positive definiteness.

$$\mathcal{L}_{modif}(X, \lambda; \nu) = F(X) - \lambda^T C(X) + \frac{1}{2\mu} ||C(X)||^2 \tag{4.22}$$

for some $0 < \nu < \mu^*$, where μ^* is chosen such that the Hessian of the modified Lagrangian is positive definite. We could now choose the matrix $\mathbf{W}_k$ to be $\nabla^2_{XX} \mathcal{L}_{modif}$ or some quasi-Newton approximation $\mathbf{B}_k$ to this matrix.

The main difficulty here is the choice of μ^*, which depends on quantities which are not normally known (e.g., bounds on the second derivatives of the problem functions).

To ensure that the **SQP** method converges from remote starting points a **merit function** Φ is employed. This function Φ is used:

- to control the size of the steps (in line search methods)
- to determine whether a step is acceptable or whether the trust-region radius needs to be modified (in trust-region methods)

It plays the role of the objective function in unconstrained optimization since we insist that each step provides a sufficient reduction in it.

The most employed are the l_1 merit function and the Fletcher's merit function.

The l_1 merit function is defined as:

$$\Phi_1(X; \mu) = F(X) + \frac{1}{\mu} ||C(X)||_1 \tag{4.23}$$

The Fletcher's merit function has the formula:

$$\Phi_1(X; \mu) = F(X) - \lambda^T C(X) + \frac{1}{2\mu} \sum C_i(X)^2 \tag{4.24}$$

The majority of line-search algorithms assume that the iteration step is obtained by means of (4.18). Other variants of **SQP** such as reduced-Hessian methods and trust region approaches compute the search direction differently.

Reduced-Hessian quasi-Newton methods are designed for solving problems in which second derivatives are difficult to compute, and for which the number of degrees of freedom in the problem, $(n - m)$, is small.

This approach is employed if we want to approximate only the reduced Hessian of the Lagrangian $Z_k^T W_k Z_k$, where Z_k is a matrix which spans the range of A_k. The update is M_k, an $(n - m) \times (n - m)$ version of the reduced-Hessian approximation.

As $(n - m)$ is small, M_k will be of high quality and the line-search computation is inexpensive. Also the reduced Hessian is much more likely to be positive definite, even when the current iterate is some distance from the solution, so that the safeguarding mechanism in the quasi-Newton update will be required less often in line search implementation.

For the trust region approach a modified model is considered:

$$\min_p \frac{1}{2} p^T W_k p + \nabla F_k^T p \tag{4.25}$$

$$\text{subject to } A_k p + C_k = 0 \tag{4.26}$$

$$||p|| \leq \Delta_k \tag{4.27}$$

The trust region radius Δ_k will be updated depending on how the predicted reduction in the merit function compares to actual reduction. If there is good agreement, the trust-region radius is unaltered or increased, whereas the radius is decreased if the agreement is poor.

Although we can simply increase Δ_k until the set of steps p satisfying the linear constraints (4.26) intersect the trust region, this approach is likely not to resolve the conflict between satisfying the linear constraints (4.26) and the trust-region constraint (4.27). A more appropriate viewpoint is to improve the feasibility of these constraints at each step and to satisfy them exactly only in the limit.

CHAPTER 5

REGULARIZATION OF ILL-POSED PROBLEMS

The computation of solutions to optimal control problems is ill-conditioned in many cases. That is, relatively large variations of parameter values are allowed for small variations of constraints and/or objective values.

The primary difficulty with ill-posed problems is that they are practically undetermined due to the condition number of the numerical implementation. Hence it is necessary to incorporate further information about the desired solution in order to stabilize the problem and to single out a useful and stable solution.

The aim of regularization is to make the computation better conditioned while changing the value of the objective only slightly. The numerical solution of the optimal control problem is obtained by minimizing a cost functional which describe the objective

$$\min_{u \in \mathcal{U}_{ad}} \boldsymbol{\mathcal{J}}(u) \tag{5.1}$$

where u is the control variable, $\mathcal{U}_{ad}$ is the set of admissible controls and $\boldsymbol{\mathcal{J}}$ is the cost functional.

Following Hansen [94], the dominant approach to regularization is to allow a certain residual to be associated with the regularized solution, with residual norm $\rho(u)$, and then use one of the following schemes:

1. Minimize $\rho(u)$ subject to the constraint that u belongs to a specified subset of $\mathcal{U}_{ad}$
2. Minimize $\rho(u)$ subject to the constraint that a measure $\omega(u)$ of the "size" of u is less than some specified upper bound δ, i.e., $\omega(u) \leq \delta$
3. Minimize $\omega(u)$ subject to the constraint $\rho(u) \leq \alpha$
4. Minimize a linear combination of $(\rho(u))^2$ and $(\omega(u))^2$

   $$\min \left\{ (\rho(u))^2 + \lambda^2 (\omega(u))^2 \right\} \tag{5.2}$$

   where λ is a specified weighting factor.

Here α, δ and λ are known as regularization terms which have to be determined and the function ω is sometimes referred to as the *smoothing norm.* The fourth scheme is the well-known Tikhonov regularization scheme [187].

Let us consider that our discrete ill-posed problem has the form

$$\min ||\mathbf{A}u - \mathbf{b}||_2 \tag{5.3}$$

where $\mathbf{A}$ is a matrix $m \times n$ $(m \geq n)$ which is ill-conditioned in the sense that all its singular values decay gradually to zero, with no gap anywhere in the spectrum.

Typically the term $\mathbf{b}$ may contain noise due to measurement and/or approximation error. This noise, in combination with the ill-conditioning of $\mathbf{A}$, means that the exact solution of (5.3) has little relationship to the noise-free solution. Instead, a regularization method is employed to determine a solution that approximates the noise-free solution. The regularization method replaces the original operator by a better-conditioned but related one. Sometimes the regularized solution is too large to solve exactly. In that case an approximate solution is computed by projection onto a smaller dimensional space, perhaps via iterative methods based on Krylov subspaces.

The conditioning of the new problem is controlled by one or more regularization parameters specific to the method. A large regularization parameter yields a new well-conditioned problem, but its solution may be far from the noise-free solution since the new operator is a poor approximation to $\mathbf{A}$. A small regularization parameter generally yields a solution very close to the noise-contaminated solution of (5.3) and hence its distance from the noise-free solution also can be large. Thus a key issue in regularization methods is to choose a regularization parameter that balances the error due to the noise with the error due to regularization.

For problems small enough that a singular value decomposition of $\mathbf{A}$ can be computed, there are well-studied techniques for computing a good regularization parameter. These techniques include the Discrepancy Principle, Generalized Cross-Validation and the $\mathcal{L}$-curve.

For larger problems treated the parameter choice is much less understood. Standard regularization methods for such a case include Tikhonov regularization or the truncated singular value decomposition. If regularization is applied to the projected problem generated by the iterative method we have an extra regularization parameter, controlling the number of iteration taken. This introduces the possibility that the standard regularization parameter that is correct for the (large) discretized problem may not be the optimal one for the lower-dimensional problem actually solved by the iteration. But the extra work due to the possible difference between the regularization parameters is offset by the fact that, in fact, we are regularizing a lower dimensional problem after projection by the iterative method.

5.1 Tikhonov regularization

One of the most common methods or regularization is Tikhonov regularization (Tikhonov and Arsenin [187]). In this method the problem (5.3) is replaced by

$$\min \left(||\mathbf{A}u - \mathbf{b}||_2^2 + \lambda^2 ||\mathbf{L}u||_2^2\right) \tag{5.4}$$

where $\mathbf{L}$ denotes a matrix, often chosen to be the identity matrix $\mathbf{I}$, a diagonal weighting matrix or a discrete derivative operator and λ is a positive scalar regularization parameter.

If an *a priori* estimate $\tilde{u}^{ap}$ of the desired regularized solution is available, then this information can be taken into account by including $\tilde{u}^{ap}$ in the discrete smoothing norm

$$\min ||\mathbf{A}u - \mathbf{b}||_2^2 + \lambda^2 ||\mathbf{L}(u - \tilde{u}^{ap})||_2^2 \tag{5.5}$$

The Tikhonov problem (5.4) has alternative formulations

$$(\mathbf{A}^*\mathbf{A} + \lambda^2 \mathbf{L}^*\mathbf{L})u = \mathbf{A}^*\mathbf{b} \tag{5.6}$$

and

$$\min \left|\left| \begin{pmatrix} \mathbf{A} \\ \lambda \mathbf{L} \end{pmatrix} u - \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix} \right|\right|_2 \tag{5.7}$$

Underlying the formulation in (5.4) is the assumption that the errors in the right-hand side are uncorrelated and with covariance matrix $\sigma_0^2 I_m$. If the covariance matrix is of more general form $\mathbf{C}\mathbf{C}^T$, where $\mathbf{C}$ has full rank m, then one should scale the least square residual with $\mathbf{C}^{-1}$ and solve the scaled problem

$$\min ||\mathbf{C}^{-1}(\mathbf{A}u - \mathbf{b})||_2^2 + \lambda^2 ||\mathbf{L}u||_2^2 \tag{5.8}$$

The most efficient and numerically stable way to compute the solution to the Tikhonov problem in (5.4) is the bidiagonalization algorithm (Elden [52]). First we want to transform the general-form problem (5.5) into the following standard-form problem

$$\min ||\bar{\mathbf{A}}\bar{u} - \bar{\mathbf{b}}||_2^2 + \lambda^2 ||(\bar{u} - \bar{u}^*)||_2^2 \tag{5.9}$$

The standard-form quantities $\bar{\mathbf{A}}, \bar{u}^*$ and $\bar{\mathbf{b}}$ take the form

$$\bar{\mathbf{A}} = \mathbf{A}\mathbf{L}_{\mathbf{A}}^{\dagger} \qquad \bar{\mathbf{b}} = \mathbf{b} - \mathbf{A}u^{\triangle} \qquad \bar{u}^* = \mathbf{L}\tilde{u}^{ap} \tag{5.10}$$

where $\mathbf{L}_{\mathbf{A}}^{\dagger}$ is the $\mathbf{A}$-weighted generalized inverse of $\mathbf{L}$.

$\mathbf{L}_{\mathbf{A}}^{\dagger}$ is defined using $\mathbf{L}^{\dagger}$, the pseudo-inverse of $\mathbf{L}$:

$$\mathbf{L}_{\mathbf{A}}^{\dagger} = \Big(\mathbf{I}_n - \Big(\mathbf{A}(\mathbf{I}_n - \mathbf{L}_{\mathbf{A}}^{\dagger}\mathbf{L})\Big)^{\dagger} \mathbf{A}\Big)\mathbf{L}^{\dagger} \tag{5.11}$$

and $u^{\triangle}$ is the unregularized component of u which is not affected by the regularization scheme:

$$u^{\triangle} \equiv \Big(\mathbf{A}(\mathbf{I}_n - \mathbf{L}_{\mathbf{A}}^{\dagger}\mathbf{L})\Big)^{\dagger} \mathbf{b} \tag{5.12}$$

The standard-form problem is then treated as a least squares problem of the form

$$\min \left|\left| \begin{pmatrix} \bar{\mathbf{A}} \\ \lambda \mathbf{I}_m \end{pmatrix} \bar{u} - \begin{pmatrix} \bar{\mathbf{b}} \\ \lambda \tilde{u}^{ap} \end{pmatrix} \right|\right| \tag{5.13}$$

This problem can be reduced to an equivalent sparse and highly structured problem. The key idea is to transform $\bar{\mathbf{A}}$ into a $m \times m$ upper bidiagonal matrix $\bar{\mathbf{B}}$ by means of alternating left and right orthogonal transformations

$$\bar{\mathbf{A}} = \bar{\mathbf{U}}\bar{\mathbf{B}}\bar{\mathbf{V}}^T \tag{5.14}$$

Software for performing the bidiagonal reduction is available in many mathematical libraries (LAPACK, LINPACK, NAG, Numerical Recipes).

Once $\bar{\mathbf{A}}$ has been reduced to a bidiagonal matrix $\bar{\mathbf{B}}$ we make the substitution $\bar{u} = \bar{\mathbf{V}}\bar{y}$ and obtain the problem

$$\min \left|\left| \begin{pmatrix} \bar{\mathbf{B}} \\ \lambda \mathbf{I}_m \end{pmatrix} \bar{y} - \begin{pmatrix} \bar{\mathbf{U}}^T \bar{\mathbf{b}} \\ \lambda \bar{\mathbf{V}}^T \tilde{u}^{ap} \end{pmatrix} \right|\right| \tag{5.15}$$

which can be solved for $\bar{y}$ in only $\mathcal{O}(m)$ operations.

A fundamental observation regarding Tikhonov regularization is that the ill-conditioning of $\mathbf{A}$ is circumvented by introducing a new problem with a new well-conditioned coefficient matrix with full rank. A different way to treat the ill-conditioning of $\mathbf{A}$ is to derive a new problem with a well-conditioned *rank-deficient* coefficient matrix. This is the philosophy behind methods based on singular value decomposition (**SVD**): truncated **SVD**, modified **SVD** and generalized **SVD**. This technique is computationally more expensive than the above approach using $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$, but it provides much more insight into the regularization problem.

5.2 Singular value decomposition

Let us remember that $\mathbf{A}$ is a $m \times n$ rectangular matrix, with $m \geq n$. The singular value decomposition (**SVD**) of $\mathbf{A}$ is a decomposition of the form

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^{n} U_i \sigma_i V_i^T \tag{5.16}$$

where $\mathbf{U} = (U_1, \dots, U_n) \in R^{m \times n}$ and $\mathbf{V} = (V_1, \dots, V_n) \in R^{n \times n}$ are matrices with orthonormal columns, $\mathbf{U}\mathbf{U}^T = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$ and where the diagonal matrix $\Sigma = diag(\sigma_1, \dots, \sigma_n)$ has nonnegative diagonal elements appearing in non-increasing order such that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \tag{5.17}$$

The numbers σ_i are called *the singular values* of $\mathbf{A}$ while the vectors U_i and V_i are the left and right singular vectors of $\mathbf{A}$, respectively.

Discrete ill-posed problems are very often characterized by the following two features of the **SVD**:

- The singular values σ_i decay gradually to zero with no particular gap in the spectrum. An increase of the dimensions of $\mathbf{A}$ will increase the number of small singular values.
- The left and right singular vectors U_i and V_i tend to have more sign changes in their elements as the index i increases, i.e., as σ_i decreases.

To see how the **SVD** gives insight into the ill-conditioning of $\mathbf{A}$, consider the following relations which follows directly from (5.16):

$$\begin{aligned} \mathbf{A}V_i &= \sigma_i U_i \qquad ||\mathbf{A}V_i||_2 = \sigma_i \\ \mathbf{A}^T U_i &= \sigma_i V_i \qquad ||\mathbf{A}U_i||_2 = \sigma_i \end{aligned}$$

If a singular value σ_i is small compared to $\sigma_1 = ||\mathbf{A}||_2$, that means that there exists a certain linear combination of the columns of $\mathbf{A}$, characterized by the elements of the right singular vector V_i such that $||\mathbf{A}V_i||_2 = \sigma_i$ is small. The same holds true for U_i and the rows of $\mathbf{A}$. In other words, a situation with one or more small σ_i implies that $\mathbf{A}$ is nearly rank deficient and the vectors U_i and V_i associated with the small σ_i are the numerical null vectors of $\mathbf{A}^T$ and $\mathbf{A}$, respectively.

Another use of the **SVD** is for the solution of the least squares problem $||\mathbf{A}x - \mathbf{b}||_2$. We can write x and respectively $\mathbf{A}x$ using the **SVD** vectors of $\mathbf{A}$:

$$x = \sum_{i=1}^{n}(V_i^T x)V_i \qquad \mathbf{A}x = \sum_{i=1}^{n}\sigma_i(V_i^T x)U_i \tag{5.18}$$

If $\mathbf{A}$ is invertible, then its inverse is given by

$$\mathbf{A}^{-1} = \sum_{i=1}^{n} V_i \sigma_i^{-1} U_i^T \tag{5.19}$$

and therefore the solution to $\mathbf{A}x = \mathbf{b}$ is

$$x = \sum_{i=1}^{n} \sigma_i^{-1}(U_i^T \mathbf{b})U_i \tag{5.20}$$

Otherwise we define the generalized inverse (Golub and Van Loan [74]) $\mathbf{A}^\dagger$ as

$$\mathbf{A}^\dagger \equiv \sum_{i=1}^{rank(\mathbf{A})} V_i \sigma_i^{-1} U_i^T \tag{5.21}$$

Then the *least squares solution* x_{LS} to the least squares problem $||\mathbf{A}x - \mathbf{b}||_2$ is given by

$$x_{LS} = \mathbf{A}^\dagger \mathbf{b} = \sum_{i=1}^{rank(\mathbf{A})} \frac{U_i^T \mathbf{b}}{\sigma_i} V_i \tag{5.22}$$

The classical algorithm for computing the **SVD** of a dense matrix is due to Golub, Kahan and Reinsch [74]. It consists of two main stages. In the first stage, $\mathbf{A}$ is transformed into upper bidiagonal form $\mathbf{B}$ by means of a finite sequence of alternating left and right Householder transformations. In the second (iterative) stage, the shifted QR algorithm is applied implicitly to the matrix $\mathbf{B}^T\mathbf{B}$ and consequently $\mathbf{B}$ converges to Σ. The left and right orthogonal transformations, if accumulated, produce the matrices $\mathbf{U}$ and $\mathbf{V}$.

This algorithm, as well as other methods for computing the **SVD** of a dense rectangular matrix, is available in many mathematical software libraries: IMSL, LAPACK, NAG, Numerical Recipes. There are also few subroutines available for large sparse matrices in the packages LANCZOS and SVDPACK.

If we consider the regularization matrix $\mathbf{L} \in R^{p \times n}$ with $m \geq n \geq p$ then we introduce the generalized singular value decomposition (**GSVD**) of the matrix pair $(\mathbf{A}, \mathbf{L})$. The generalized singular values of $(\mathbf{A}, \mathbf{L})$ are essentially the square roots of the generalized eigenvalues of the matrix pair $(\mathbf{A}^T\mathbf{A}, \mathbf{L}^T\mathbf{L})$.

Assuming that $\mathbf{L}$ has full row rank the **GSVD** is a decomposition of $\mathbf{A}$ and $\mathbf{L}$ in the form

$$\mathbf{A} = \mathbf{U} \begin{pmatrix} \Sigma & 0 \\ 0 & \mathbf{I}_{n-p} \end{pmatrix} \mathbf{Z}^{-1} \qquad \mathbf{L} = \mathbf{V}(\mathbf{M}, \mathbf{0})\mathbf{Z}^{-1} \tag{5.23}$$

The columns of $\mathbf{U} \in R^{m\times n}$ and $\mathbf{V} \in R^{p\times p}$ are orthonormal, $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I}_p$. $\mathbf{Z} \in R^{n\times n}$ is nonsingular with columns that are $\mathbf{A}^T\mathbf{A}$ orthogonal. $\Sigma = diag(\sigma_1, \ldots, \sigma_p)$ and $\mathbf{M} = diag(\nu_1, \ldots, \nu_p)$ are $p \times p$ diagonal matrices. The diagonal elements are nonnegative, ordered such that

$$0 \le \sigma_1 \le \cdots \le \sigma_p \le 1 \qquad 1 \ge \nu_p \ge \cdots \ge \nu_1 \ge 0 \tag{5.24}$$

and normalized such that

$$\sigma_i^2 + \nu_i^2 = 1 \tag{5.25}$$

Then the generalized singular values γ_i of $(\mathbf{A}, \mathbf{L})$ are defined as the ratios

$$\gamma_i = \frac{\sigma_i}{\nu_i} \tag{5.26}$$

Since

$$\mathbf{Z}^T\mathbf{A}^T\mathbf{A}\mathbf{Z} = \begin{pmatrix} \Sigma^2 & 0 \\ 0 & \mathbf{I}_{n-p} \end{pmatrix} \qquad \mathbf{Z}^T\mathbf{L}^T\mathbf{L}\mathbf{Z} = \begin{pmatrix} \mathbf{M}^2 & 0 \\ 0 & \mathbf{0} \end{pmatrix} \tag{5.27}$$

we can see that (γ_i^2, Z_i) are the generalized eigensolutions of the pair $(\mathbf{A}^T\mathbf{A}, \mathbf{L}^T\mathbf{L})$ associated with p finite generalized eigenvalues.

The following three characteristic features of **GSVD** are common for a discrete ill-posed problem :

- The generalized singular values γ_i decay gradually to zero with no gap in the spectrum. An increase of the dimensions of $\mathbf{A}$ will increase the number of small generalized singular values.
- The singular vectors U_i, V_i and Z_i tend to have more sign changes in their elements as the corresponding γ_i decreases.
- If $\mathbf{L}$ approximates a derivative operator, then the last $n - p$ columns Z_i of $\mathbf{Z}$ have very few sign changes, since they are the null vectors of $\mathbf{L}$.

5.3 Hybrid methods: projection plus regularization

If the problem is too large one may consider regularization achieved through projection onto a subspace (e.g., Fleming [56]). The truncated **SVD** (**TSVD**) is an example of such projection: the solution is constrained to lie in the subspace spanned by the singular vectors corresponding to the largest $n - l$ singular values, where l is the number of terms to be dropped from the sum.

Hybrids methods were introduced by O'Leary and Simmons [153]. These methods combine a projection method with a direct regularization method like **TSVD** or Tikhonov regularization. The problem is projected onto a particular subspace of dimension k, but typically the restricted operator is still ill-conditioned. A regularization method is applied to the projected problem. Since the dimension k is usually small relative to n, regularization of the restricted problem is much less expensive. Yet, with an appropriately chosen subspace, the end results can be very similar to those achieved by applying the same direct regularization technique to the original problem (Kilmer and O'Leary [120]). Because the projected problems are usually generated iteratively by a Lanczos method, this approach is useful when the matrix is sparse or structured in such a way that matrix-vector products can be handled efficiently with minimal storage.

5.4 Parameter selection methods

No regularization method is complete without an algorithm for choosing the regularization parameter. We discuss here several parameter-choice methods.

Without loss of generality restrict our discussion to the standard form case. This is possible due to the relations

$$||\mathbf{L}u||_2 = ||\mathbf{u}||_2 \qquad ||\mathbf{A}u - \mathbf{b}||_2 = ||\bar{\mathbf{A}}\bar{u} - \bar{\mathbf{b}}||_2 \tag{5.28}$$

which ensure that application of a norm-based parameter-choice rule to the original problem with $\mathbf{A}$ and $\mathbf{b}$, or to the standard-form problem with $\bar{\mathbf{A}}$ and $\bar{\mathbf{b}}$, yields exactly the same regularization parameter.

We consider the norm of the error in the right hand side

$$||e||_2 = ||\mathbf{b} - \mathbf{b}^{exact}||_2 \tag{5.29}$$

Parameter-choice methods can be divided into two classes depending on their assumptions about the error norm $||e||_2$:

1. Methods based on knowledge, or a good estimate, of $||e||_2$ (e.g., the Discrepancy Principle).

2. Methods that do not require $||e||_2$, but instead seek to extract this information from the given right-hand side (e.g., the Generalized Cross Validation and L-curve Criterion).

The most widespread $||e||_2$-based method is the Discrepancy Principle **DP** (Morozov [146]). We suppose that the ill-posed problem is consistent in the sense that $\mathbf{A}u^{exact} = \mathbf{b}^{exact}$ holds exactly. Under **DP**, we consider all solutions with $||\mathbf{A}u - \mathbf{b}||_2 \leq \delta_\epsilon$ and select from these solutions the one that minimizes the norm of u. This can be written as:

$$\text{Minimize } ||u||$$
$$\text{subject to } ||\mathbf{A}u - \mathbf{b}||_2 \leq \delta_\epsilon$$

Generalized cross-validation **GCV** (Golub et al. [73], Wahba [199], [200]) is based on statistical considerations, namely that a good value of the regularization parameter should predict missing data values. The main idea is to find a parameter λ that minimizes the **GCV** functional

$$\mathbf{G}(\lambda) = \frac{||(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^{\oplus})\mathbf{b}||_2}{(trace(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^{\oplus}))^2} \tag{5.30}$$

where $\mathbf{A}_\lambda^{\oplus}$ denotes the matrix that maps the right hand side $\mathbf{b}$ onto the regularized solution x_λ. **GCV** chooses aregularization parameter that is not too dependent on any one data measurement.

The final parameter-choice method discussed here is the $\mathcal{L}$-curve criterion (**LCC**). The $\mathcal{L}$-curve is defined as a parametric plot of the norm of the regularized solution $||\mathbf{L}u_{reg}||_2$ versus the corresponding residual norm $||\mathbf{A}u_{reg} - \mathbf{b}||_2$, with the regularization parameter λ as the parameter. As the regularization parameter increases the norm of the solution decreases while the residual increases.

The best regularization parameter should lie on the corner of the $\mathcal{L}$-curve. For values higher than this the residual increases without reducing the norm of the solution much, while for values smaller than this, the norm of the solution increases rapidly without much decrease in residual.

In practice only a few points on the $\mathcal{L}$-curve are computed and the corner is located by approximate methods, estimating the point of a maximum curvature (Hansen and O'Leary [95]).

CHAPTER 6

DESCRIPTION OF THE PHYSICAL PHENOMENA FOR THE FLOW AROUND A CYLINDER

When a fluid flows past a stationary body or, equivalently, when a body moves in a fluid at rest, a region of disturbed flow is always formed around the body. The extent of the disturbed flow region is largely dependent on the shape, orientation and size of the body, the velocity and viscosity of the fluid and may be influenced by a wide variety of small disturbances.

A particularly large and usually unsteady separated flow is generated by bluff bodies. Bluff bodies may have sharp edges on their circumferences such as flat plates, triangular, rectangular and polygonal cylinders or may be rounded like circular, elliptical and arbitrary oval cylinders. The common feature of flows around bluff bodies is the development of similar flow structures in the separated region.

Experiments showed the division of the disturbed flow field into four regions:

1. One narrow region of retarded flow

2. Two boundary layers attached to the surface of the cylinder

3. Two side-wise regions of displaced and accelerated flow

4. One wide downstream region of separated flow called the *wake*

The upstream retarded flow region presents high fluctuations in velocity. The inherently unstable retarded flow forms unsteady flow structures in a streamwise direction.

The boundary layers around the cylinder are subject to a favorable pressure gradient followed by a small region of adverse pressure gradient before separation. The separated boundary layers continue to develop downstream as free shear layers and they initially border the near-wake.

In the third region, the displaced flow is vigorously entrained by the low pressure in the wake. The extent of the displaced region is strongly affected by the vicinity of confining walls of wind or water tunnels, phenomenom known as the *blockage* effect.

Large flow structures are formed in the near wake and gradually decay along the wake. The formation and decay of the flow structures depend on the state of flow which may be laminar, transitional or turbulent.

Reynolds (1883) discovered that transition from laminar to turbulent flow in a smooth pipe depends upon the fluid density ρ, the viscosity μ, the velocity V and the internal diameter of the pipe d. This transition takes place within a range of the Reynolds number $Re = \frac{\rho V d}{\mu}$.

The state of flow may be fully laminar L, it may be in any of the three transitions TrW, TrSL and TrBL, or, respectively, fully turbulent T.

6.1 The laminar state of flow

The laminar state of the disturbed flow can be subdivided into three basic flow regimes:

1. Non-separation regime: $0 < Re < (4 - 5)$
2. Closed near-wake regime: $(4 - 5) < Re < (30 - 48)$
3. Periodic laminar regime: $(30 - 48) < Re < (180 - 200)$

The flow in the first region is firmly attached to the surface of the cylinder all around the circumference. The trail of steady and symmetric laminar shear layers does not form a visible wake in the non-separation regime.

Separation initiates at $Re = 4$ to 5 when a distinct, steady, symmetric and closed *near-wake* is formed. The free shear layers meet at the end of the near-wake at the confluence point.

The elongated closed near-wake becomes unstable for $Re > (30 - 48)$ and a sinusoidal oscillation of shear layers commences at the confluence point. The amplitude of the trail oscillation increases with rising Re. The final product is a staggered array of laminar eddies.

Benard (1908) was the first to sketch the alternate procession of eddies behind a towed circular cylinder in water based on visible dimples on the water surface. Von Karman (1911) considered the stability of two rows of vortices theoretically and stimulated a widespread interest. The alternating eddies develop gradually along the laminar wake. Taneda (1956) proposed a subdivision of the periodic laminar regime into two separate phases: oscillating free shear layers without eddies and a Karman vortex street formed behind the closed near wake. The *Karman vortices* play an important role in the formulation of the optimal control problem for flow around a rotating cylinder, since our goal was to suppress the formation of these vortices using the rotation rate of the cylinder as the control parameter.

6.2 The transition states of flow

Dryden (1941) first noted the succession of transitions with Re in various regions of the disturbed flow. Experiments showed the development of transitions in three disturbed regions: wake (TrW), shear layers (TrSL) and boundary layers (TrBL).

The first transition TrW occurs in the wake and it was discovered by Roshko (1954) in the range of Re being one order of magnitude lower than in pipe flow experiments. The second transition TrSL appears in the free shear layers. It was first noted by Linke (1931) then examined in detail by Bloor (1964) and Gerrard(1965). The third transition reaches the boundary layers at separation. It was discovered by Wieselberger (1914) and Prandtl (1916).

The laminar periodic wake becomes unstable at higher Re farther downstream in the wake. Gradually transition spreads upstream with increasing Re until the eddy becomes turbulent during its formation.

Transition-in-wake state can be divided into two regimes:

- TrW1: Transition of laminar eddies in the wake for $$(180 - 200) < Re < (220 - 250)$$
- TrW2: Transition of an irregular eddy during its formation for $$(220 - 250) < Re < (350 - 400)$$

Between the two regimes TrW1 and TrW2 the laminar wake instability mode of eddy formation and shedding is replaced by the turbulent eddy roll up and shedding mode from the cylinder. The change of the eddy shedding mode is reflected by the different variation in shedding frequency expressed through a non-dimensional Strouhal number St defined by $St = \dfrac{f_K D}{U_0}$, where f_K is the Karman vortex street frequency and D is the diameter of the cylinder.

The second transition TrSL takes place along the free shear layers while the boundary layers remain fully laminar. There are three phases of transition along the free-shear layers:

- TrSL1: Development of transition waves for $$(350 - 400) < Re < (10^3 - 2 \times 10^3)$$
- TrSL2: Formation of transition eddies for $$(10^3 - 2 \times 10^3) < Re < (2 \times 10^4 - 4 \times 10^4)$$
- TrSL3: Burst to turbulence for $$(2 \times 10^4 - 4 \times 10^4) < Re < (10^5 - 2 \times 10^5)$$

The transition waves appear first as undulations of the free shear layers. As Re increases the transition waves roll up into discrete eddies, along the free shear layer, before becoming turbulent and then roll up in alternate eddies. Finally a sudden burst to turbulence occurs in the free shear layers near the side of the cylinder and the formation of eddies takes place close to the rear of the cylinder.

Five regimes were suggested for the transition-in-boundary-layers TrBL (Zdravkovich [209]):

- TrBL0: Precritical regime for $(10^5 - 2 \times 10^5) < Re < (3 \times 10^5 - 3.4 \times 10^5)$
- TrBL1: One-bubble regime for $(3 \times 10^5 - 3.4 \times 10^5) < Re < (3.8 \times 10^5 - 4 \times 10^5)$
- TrBL2: Two-bubble regime for $(3.8 \times 10^5 - 4 \times 10^5) < Re < (5 \times 10^5 - 10^6)$
- TrBL3: Supercritical regime for $(5 \times 10^5 - 10^6) < Re < (3.4 \times 10^6 - 6 \times 10^6)$
- TrBL4: Post-critical regime for $(3.4 \times 10^6 - 6 \times 10^6) < Re < (unknown)$

The precritical regime is characterized by the first onset of transition in free shear layers along separation lines. There is an initial fall in the drag coefficient while the eddy shedding remains prominent. TrBL0 terminates abruptly at certain *Re* with a discontinuous fall in the drag coefficient and with a jump in the frequency of eddy shedding.

For the next regime, TrBL1, the pressure distribution is asymmetric. On one side of the cylinder the free shear layers underwent sufficient transition to be able to reattach onto the cylinder surface. The closed thin separated region was termed a *separation bubble.* The asymmetric single-bubble regime TrBL1 terminates at higher *Re* with yet another discontinuous fall in the drag and a jump in the shedding frequency when a second bubble is formed on the other side of the cylinder.

The symmetric two-bubble regime TrBL2 represents an intricate combination of laminar separation transition, reattachment and turbulent separation on the boundary layers on both sides of the cylinder. Both TrBL1 and TrBL2 are very sensitive to disturbances and can be eliminated by a sufficiently rough surface and/or turbulent free stream.

Further increase in *Re* brings transition to the primary laminar separation line in an irregular manner. This leads to the disruption and fragmentation of separation bubbles along the span of the cylinder. The irregularly fragmented separation lines prevent periodic eddy shedding, which is the main feature of the super-critical regime TrBL3.

Roshko (1961) discovered that eddy shedding reappears at higher *Re* when the boundary layers are turbulent before separation all along the span. This regime, TrBL4, is characterized by the transition in boundary layers being somewhere between the stagnation and separation lines. As *Re* increases, the transition advances asymptotically towards the stagnation line and hence the value of *Re* for the upper end of TrBL4 is hard to define.

6.3 Fully turbulent state of flow

This state of flow is reached when all disturbed flow regions around the cylinder are turbulent. It is not known at present at which value of *Re* the turbulent state starts.

The flow past the cylinder and the associated drag and eddy shedding are expected to be invariant provided that the influencing parameters are kept small. However this becomes hardly possible because compressibility effects in air and cavitation in water cannot be avoided at very high *Re* and they become governing parameters.

6.4 Evolution of the fluid-dynamic section

The flow structures described in the previous sections determine the magnitude, direction and time variation of the fluid-dynamic force exerted upon the cylinder. For example, the symmetric flow regimes L1 and L2 in the laminar state give rise to a steady resistance, while the laminar periodic regime L3 generates a regular periodic force with components in both drag and lift direction (drag and lift forces represent the resultant force along and respectively normal to the free stream velocity).

The fluctuating drag and lift forces are denoted by $C_D^{'}$ and $C_L^{'}$ and the time-averaged values by C_D and C_L. The drag force C_D is produced by viscous friction along the surface C_{D_f} and by an asymmetric pressure distribution on the upstream and downstream side of the cylinder C_{D_P}:

$$C_D = C_{D_f} + C_{D_P} \tag{6.1}$$

The viscous friction C_{D_f} is significant in the laminar state but becomes negligible beyond the end of TrSL state of flow. The variation of pressure-drag C_{D_P} is closely related to the flow regimes. It oscillates, with three local minimum values corresponding to the elongated, steady and closed near-wake at the end of L2, the longest length of eddy formation region between TrSL1 and TrSL2 and the separation bubbles on both sides of the cylinder in TrBL2 respectively.

The fall in C_D and the appearance of mean C_L occur at the beginning of the single-bubble regime. TrBL1 is followed by another fall in C_D and C_L at the start of the two-bubble regime TrBL2.

The fluctuating lift $C_L^{'}$ is always greater than the fluctuating drag $C_D^{'}$. The latter has two components: $C_{DS}^{'}$ which is sinusoidal and $C_{DT}^{'}$ which is random and produced by turbulence. C_L has also two similar components $C_{LS}^{'}$ and $C_{LT}^{'}$ (except in the L3 regime). $C_{LS}^{'}$ is dominant in TrW2 and TrSL3 and vanishes in TrBL3. In the post-critical state TrBL4 $C_{LS}^{'}$ has the same order of magnitude as $C_{LT}^{'}$.

6.5 Additional considerations for the flow corresponding to the Reynolds number in the range $0 < Re < 1000$

For our research we considered the flow around a cylinder for the Reynolds number in the interval $[40, 1000]$. A more *in-depth* analysis of the flow in the range considered provides better insight for the validity of the optimized numerical results, by relating the physical phenomena to the corresponding numerical values obtained during and after the process of flow optimization.

The flow at very low Reynolds number Re is dominated by viscous forces to such an extent that all disturbed regions remain laminar. The separation appears for $Re \approx 5$. The most notable feature of the regime in the range $5 < Re < 40$ is a steady separated region in the form of a laminar closed near-wake behind the cylinder.

At $Re = 3.5$ it was observed that the cylinder is “pushing” and “dragging” thick shear layers by the action of large viscous forces. This “pushing” and “dragging” becomes self-evident by towing a cylinder through a fluid at rest. These two actions produce a large resistance force. The sidewise and upstream displacement of fluid from cylinder can be strongly influenced by the vicinity of side walls (it was observed that a cylinder confined in a 500D wide container was still affected by the side walls at very small Reynolds numbers).

This influence, called the wall blockage, occurs in most experiments and it is not present in applications. The confining walls of wind and water tunnels restrict the disturbed flow sidewise and impose an additional pressure gradient. The blockage ratio $\frac{G}{D}$ (where G is the distance between the walls) is enhanced by thick boundary layers and it may become the dominant parameter.

The magnitude of viscous forces decreases with increasing Re until separation occurs at a certain Re_{sep}. The separation was first observed by using smoke visualization. The blockage has a strong effect on Re_{sep}. It is difficult to determine Re_{sep} experimentally because the size of the near-wake is small and separation occurs in a region where the velocity is also very small. The appearance of a steady separated region confined in a closed and symmetric near-wake is marked by a noticeable change in pressure distribution. The adverse pressure gradient is relieved by separation.

The closed near-wake characteristic for $5 < Re < (30 - 48)$ is symmetric, steady and it is formed as the separated shear layers merge downstream. A metamorphosis of the near-wake was observed if Re increases from 20 to 40. There is a sequence of elongation and then obliteration of the initially closed near-wake. The formation of a new near-wake is accomplished by secondary separations of the free shear layers from the near-wake.

Another unexpected feature of the closed near-wake regime is that the streamlines displaced by the cylinder do not follow the shape of the near-wake boundaries. One may observe a widening of the streamlines instead, which increases away from the cylinder.

The steady, elongated and closed near wake becomes unstable when $Re > Re_{osc}$, where the subscript *osc* stands for *oscillation*. The transverse oscillation starts at the end of the near-wake and initiates a wave along the trail. As Re was increased from 40 to 60 the development of secondary separations of the free shear layers from the near-wake boundary is accompanied by the transverse oscillation of the trail. The secondary separations prevent the free shear layers from meeting at the confluence point as they do behind the steady and closed near-wake.

The near-wake instability initiates a wavy trail for $Re > Re_{osc}$. The wavelength of the trail gradually decreases with rising Re. At the same time the amplitude of crests and troughs of the wavy trail increases with rising Re and the free shear layers begin to roll up and form eddies.

A fully developed Karman vortex (eddy) street has three distinct features:

1. The staggered vortices are not shed from the cylinder but initiate at the end of the closed near wake;

2. The roll up is gradual and takes place along the wake until the pattern becomes "frozen";

3. The widening of the wake is accomplished by the entrainment of the external irrotational fluid.

We define the vortex shedding period (VSP) as the inverse of the Strouhal number. The plot St versus Re shows a logarithmic increase of St as we increase Re. From experimental data it was observed that a discontinuous drop in shedding frequency occurs in the range $80 < Re < 130$. This suggested the existence of cells of different shedding frequencies along the span of the cylinder. But the coexistence of different shedding frequencies along the span cannot explain the discontinuity which occurs at a certain Reynolds number Re_d. A better explanation of the phenomenon, verified experimentally, is that the discontinuity in the frequency is produced by a transition mode from one slanting shedding mode to another slanted mode.

Flow in the laminar periodic wake is two-dimensional if all eddy filaments are parallel to the cylinder axis. The flow is truly two-dimensional in the range $40 < Re < 80$. For $80 < Re < 120$ the wake is sensitive to disturbances and may become three-dimensional. The majority of experiments showed that the laminar eddy filaments were either slanted or wavy spanwise as $Re \geq 120$. This implied that the periodic wake has three-dimensional characteristics for that range of Re, although a small number of experiments obtained a two-dimensional wake even for the range $120 < Re < 180$.

Considerable effort has been devoted for discovering what causes the existence of two modes of flow in the laminar periodic wake. The eddy filaments are more or less parallel to the cylinder axis in the initial phase of flow. The slanted eddy filaments developed subsequently as the effect of the ends spread along the span. It was suggested that the slantwise shedding is an intrinsic feature of the flow which arise from a difference in the end effects, although the magnitude of the effect may depend on the particular end effects of the cylinder.

Based on these observations researchers found methods to induce parallel vortex shedding: by fitting end plates on both sides of the cylinder, by addition of two short cylinders at both ends, by placing too cylinders of large diameters perpendicular to and upstream of the model cylinder or by applying suction at both ends of the cylinder.

The eddy formation is completed when a maximum concentration of vorticity is attained. The distance of that point from the cylinder is named *the length of the eddy formation region L_f*. Beyond L_f the viscous dissipation and diffusion gradually reduce the strength of eddies. It might be expected that the decay of laminar vortices by diffusion and viscous dissipation would eventually annihilate the eddy street far downstream.

Experiments showed that after an almost complete obliteration of the primary vortex street a secondary eddy street gradually emerges in the far-wake. In some cases a tertiary eddy street followed the secondary one. There is wide agreement on the fact that the secondary eddy street can be found for $100 < Re < 160$. For the range $70 < Re < 100$ there were results showing the secondary vortex street as well as research which could not detect it. After $Re > 160$ the wake becomes irregular and eventually turbulent, making the interpretation of flow visualization much more difficult.

All laminar flows eventually become unstable above a certain Re and undergo transition to turbulence. The flow in a wake does not become fully turbulent as soon as it ceases to follow the laws of laminar flow. There is a finite *transition region* characterized by the random initiation and growth of irregularities. The transition in periodic laminar wakes is further complicated by the viscous diffusion and mutual interaction.

As mentioned in the beginning of this chapter the transition-in-wake TrW may be divided into two flow regimes:

1. Lower transition regime TrW1: the vortices are formed laminar and regular, but become irregular and transitional further downstream

2. Upper transition regime TrW2: the eddies are formed laminar and irregular, but become partly turbulent before they are shed and carried downstream.

The transitional wake in TrW2 is still surrounded by laminar free shear layers. It has been shown that a two-dimensional vortex filament subjected to three-dimensional disturbances is distorted progressively by its own induction. The continuous distortion of laminar eddy filaments leads ultimately to their breakdown.

The distortions of eddy filaments appeared at randomly disposed spanwise positions and followed each other at the same spanwise position. They were called “fingers” because they point towards the cylinder. They first appeared for $Re > 150$ and persisted for 2 or 3 cycles, but as Re increased they appeared more frequently at each position and in clumps of a larger number.

The three-dimensional and random appearance of fingers may be related to a low-frequency signal detected by a hot wire. Low frequency irregularities were found for $200 < Re < 400$. They became more vigorous downstream and eventually rendered the wake turbulent.

The shedding frequency of laminar and turbulent eddies has been measured by many researchers, starting with Strouhal (1878). The following ranges were suggested:

1. *stable range*, $40 < Re < 150$: regular velocity fluctuations and rising St;

2. *unstable range*, $150 < Re < 300$: irregular bursts in velocity fluctuations and St unstable;

3. *irregular range*, $Re > 300$: irregular and periodic, St constant.

The boundary between TrW1 and TrW2 is marked by a jump in St at $Re \approx 250$ which separates rising St from $St = const.$

There are two modes of eddy shedding: low-speed mode and high-speed mode. The distinct feature of the low-speed mode is the sinusoidal trail and gradual roll up of free shear layers at crests and troughs. For the high-speed mode the vortices are not mutually connected. The upper eddy is formed in an almost stationary position and the cutoff of the upper shear layer is executed by the lower eddy on the opposite side. Measurements have demonstrated there was no smooth transition from low-speed to high-speed mode of eddy shedding.

The TrW state of flow is associated with transition to turbulence in wake. This means that all eddies are formed laminar in TrW1 and TrW2 regimes and become turbulent downstream. Turbulent eddies are produced by mixing with the free stream flow around them. The eddies induce transverse flow across the wake which is an intrinsic feature of the high-speed mode of eddy shedding. Experiments showed that at $Re = 210$ we have fully laminar flow in both wake and eddies.

As Re increases to 270 the transverse flow between eddies becomes turbulent in the confluent region. As Re increases again to 400 the transition in confluent regions becomes more extended. Based on these observations it was suggested that the transition to turbulence in laminar eddies is initiated by the entrainment of turbulent fluid from the confluent regions into the otherwise laminar eddies, hypothesis which was confirmed experimentally.

It has been shown that the initiation of transition in TrW1 is associated with the appearance of "fingers" and the latter are always irregular and three-dimensional. This indicates that the formation of "fingers" and three-dimensional flow should be postponed, in order to suppress transition. This was achieved by several methods, such as by forcing the cylinder to oscillate at high frequency or by enhancing two-dimensionality by placing two parallel cylinders in tandem arrangement to the oncoming free stream.

The transition to turbulence in the free shear layers (TrSL) develops through distinct phases as the Reynolds number rises (Zdravkovich [209]):

- TrSL1 (lower sub-critical regime): transition waves appear along free shear layers and stabilize the near wake;

- TrSL2 (intermediate sub-critical regime): transition vortices are formed as a chain along free shear layers and they precede the transition to turbulence;

- TrSL3 (upper sub-critical regime): an immediate transition to turbulence close to the cylinder is accompanied by a very short near-wake.

Since our research considered the Reynolds number in the range $2 < Re < 1000$, we will end this chapter with considerations about TrSL1.

In the above regions turbulent eddies are regularly formed, periodically shed and rapidly dissipated along the wake. The wake energy decays rapidly as the fluid moves away from the cylinder.

The Karman vortex street evolves gradually by a roll-up of the free shear layers at crests and troughs of the wavy trail beyond $Re = 60$. Similar laminar waves are observed in boundary layers before the transition to turbulence. Transition waves, an analogous counterpart, are found in the free shear layers emanating from circular cylinders beyond $Re = 500$. The apparent similarity of all three kinds of waves suggests a universal mechanism of transition to turbulence.

We did not consider Reynolds numbers for which the flow is predominantly turbulent. For more about the characteristics of the turbulence regime for the flow around a cylinder the reader is referred to Zdravkovich [209]. An overview of turbulent flow research in the areas of simulation and modeling is provided by Gatski et al. [62].

CHAPTER 7

OPTIMAL CONTROL OF A FLOW AROUND A ROTATING CYLINDER

7.1 The governing equations of the model

Let B denote a circular cylinder enclosed by an impermeable boundary Γ, while the two-dimensional exterior domain $\mathbf{D} = \mathbf{R}^2 \setminus \{B \cup \Gamma\}$ is the region occupied by an incompressible viscous fluid (for numerical purposes, the domain will be restricted to a rectangle in $\mathbf{R}^2$).

The fluid is moving with velocity U_0 in the x-direction and the cylinder rotates counterclockwise with angular velocity $\mathbf{\Omega}$.

The problem can be mathematically described by the 2-D unsteady Navier-Stokes equations, where (u, v) is the velocity vector and p is the pressure:

$$\frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = \frac{1}{Re}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) - \frac{\partial (u^2)}{\partial x} - \frac{\partial (uv)}{\partial y} \quad \text{in } \mathbf{D} \tag{7.1}$$

$$\frac{\partial v}{\partial t} + \frac{\partial p}{\partial y} = \frac{1}{Re}\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}\right) - \frac{\partial (uv)}{\partial x} - \frac{\partial (v^2)}{\partial y} \quad \text{in } \mathbf{D} \tag{7.2}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \quad \text{in } \mathbf{D} \tag{7.3}$$

subject to initial condition

$$(u, v)|_{t=0} = (u_0, v_0) \quad \text{in } \mathbf{D}. \tag{7.4}$$

The equations are non dimensional. Re is the Reynolds number defined as $Re = \frac{U_0 d}{\nu}$, where d is the diameter of the cylinder and ν is the kinematic viscosity $\nu = \frac{\mu}{\rho}$, with μ the viscosity and ρ the density.

No-slip boundary condition are enforced at the upper and lower boundaries; an inflow boundary condition is applied at the left boundary:

$$u = U_0 \qquad \text{and} \qquad v = 0 \tag{7.5}$$

and an outflow boundary condition at the right boundary:

$$\frac{\partial u}{\partial x} = 0 \qquad \text{and} \qquad \frac{\partial v}{\partial x} = 0. \tag{7.6}$$

On the surface of the cylinder the velocity is equal to the angular velocity $\mathbf{\Omega} = (\mathbf{\Omega}_x, \mathbf{\Omega}_y)$:

$$u = \mathbf{\Omega}_x \qquad v = \mathbf{\Omega}_y. \tag{7.7}$$

7.2 Space and time discretization

The region **D** is discretized using a staggered grid as presented in Fig. 7.1 (Griebel et al. [79]). The pressure p is located at the cell centers, the horizontal velocity u at the midpoints of the vertical cell edges and the vertical velocity v at the midpoints of the horizontal cell edges. Cell (i, j) occupies the spatial region $[(i-1)\Delta x, i\Delta x] \times [(j-1)\Delta y, j\Delta y]$ and the corresponding index (i, j) is assigned to the pressure at the cell center as well as to the u-value at the right edge and the v-velocity at the upper edge of the cell.

Consequently, not all extremal grid points come to lie on the domain boundary. The vertical boundaries, for instance, carry no v-values, just as the horizontal boundaries carry no u-value. For this reason, an extra boundary strip of grid cells is introduced (see Fig. 7.2), so that the boundary conditions may be applied by averaging the nearest grid points on either side.

We require that the discretized values of u and v on the boundary cells are equal to the components of the angular velocity on the circle. This boundary condition is enforced by averaging the values on either side of the boundary and setting this average to be equal to the angular velocity value.

The continuity equation (7.3) is discretized at the center of each cell by replacing the spatial derivatives with centered differences using half of the mesh width. The momentum equation (7.1) for u, on the other hand, is discretized at the midpoints of the vertical cell edges, while the momentum equation (7.2) for v is discretized at the midpoints of the horizontal edges.

The second derivatives of u and v as well as the spatial derivatives of pressure are discretized using central differences with half the step size.

The discretization of the convective terms $\partial(u^2)/\partial x, \ldots, \partial(uv)/\partial y$, however, poses some difficulties. The first approach was to employ averages of u and/or v. For example, the discrete $\partial(uv)/\partial y$ has the formula

$$\left[\frac{\partial(uv)}{\partial y}\right]_{i,j} = \frac{1}{\Delta y}\left(\frac{(v_{i,j}+v_{i+1,j})}{2}\frac{(u_{i,j}+u_{i,j+1})}{2} - \frac{(v_{i,j-1}+v_{i+1,j-1})}{2}\frac{(u_{i,j-1}+u_{i,j})}{2}\right)$$

Because the convective terms in the momentum equation become dominant at high Reynolds numbers or high velocities, it is necessary to use a mixture of the central differences and the donor-cell discretization. The discrete $\partial(uv)/\partial y$ becomes

$$\begin{aligned}\left[\frac{\partial(uv)}{\partial y}\right]_{i,j} &= \frac{1}{\Delta y}\left(\frac{(v_{i,j}+v_{i+1,j})}{2}\frac{(u_{i,j}+u_{i,j+1})}{2} - \frac{(v_{i,j-1}+v_{i+1,j-1})}{2}\frac{(u_{i,j-1}+u_{i,j})}{2}\right)\\ &+ \gamma\frac{1}{\Delta y}\left(\frac{|v_{i,j}+v_{i+1,j}|}{2}\frac{|u_{i,j}+u_{i,j+1}|}{2} - \frac{|v_{i,j-1}+v_{i+1,j-1}|}{2}\frac{|u_{i,j-1}+u_{i,j}|}{2}\right)\end{aligned}$$

where γ is a parameter which should be chosen such that

$$\max_{i,j}\left(\left|\frac{u_{i,j}\Delta t}{\Delta x}\right|, \left|\frac{v_{i,j}\Delta t}{\Delta y}\right|\right) \leq \gamma \leq 1$$

The time discretization is explicit in the velocities and implicit in the pressure: i.e., the velocity field at each time step t_{n+1} can be computed once the corresponding pressure was computed. The time step is required to satisfy the stability condition :

$$\delta t = \tau \min\left(\frac{Re}{2}\left(\frac{1}{\Delta x^2}+\frac{1}{\Delta y^2}\right)^{-1}, \ \frac{\Delta x}{u_{\max}}, \ \frac{\Delta y}{v_{\max}}\right).$$

where $\tau \in [0, 1]$ is the Courant-Fredrichs-Levy (CFL) number (set to 0.6 in the code).

The domain is a rectangle of 22.0 units in length and 4.1 units in width. The cylinder (located inside the rectangle) measures 1.0 units in diameter and is situated at a distance of 1.5 units from the left boundary and 1.6 units from the upper boundary of the domain.

The cylinder is rotating with an angular velocity which can be either constant in time or a time-dependent function.

Figure 7.7 shows the uncontrolled flow for this domain.

7.3 Formulation of the optimal control problem

The control problem consists in finding the optimal angular velocity of the cylinder such that the Karman vortex shedding in the wake of the cylinder is suppressed.

In order to find the optimal value(s) of the angular velocity of the cylinder, we minimize a cost functional which depends on the state variables as well as on the control variables. The control variables are the rotation parameters: amplitude A and frequency F.

We define the speed ratio $\alpha \equiv \frac{a\Omega}{U}$, where a is the radius of the cylinder, Ω is the angular velocity and U is the free stream velocity.

We considered both the **constant rotation** case: $\alpha(t) = A$ as well as
the **time harmonic rotary oscillation** case: $\alpha(t) = A\sin(2\pi F t)$.

The vector of control parameters is $\mathbf{\Lambda} = A$ or $\mathbf{\Lambda} = (A, F)$ respectively.

With these notations, the optimal control problem becomes:

IF $\mathbf{\Lambda}$ IS THE VECTOR OF PARAMETERS WHICH DETERMINE THE ANGULAR VELOCITY OF THE CYLINDER, MINIMIZE THE COST FUNCTIONAL $\mathcal{J}$ WITH RESPECT TO $\mathbf{\Lambda}$ SUBJECT TO THE CONSTRAINTS IMPOSED BY THE 2-D UNSTEADY NAVIER-STOKES EQUATIONS MODEL.

Based on recent research work (e.g., Abergel and Temam [1], Burns and Ou [23], Ou [154], Ghattas and Bark [63], Berggren [14], Bewley et al. [17]), several possible approaches to control the behavior of the flow can be employed, such as:

- **flow tracking** (the velocity field should be "close" to a *desired field*);
- **enstrophy minimization** (the vorticity is minimized);
- **dissipation function** (minimize the rate at which heat is generated by deformations of the velocity field).

In this research work we considered only **flow tracking** and **vorticity minimization**. The mathematical expressions of the corresponding cost functionals are provided below.

We considered a cost functional for vorticity minimization of the form:

$$\mathcal{J}(\Lambda) = \frac{1}{2}\int_{t_1}^{t_2}\int_{\mathbf{D}}(\zeta^2)d\mathbf{D}dt \tag{7.8}$$

where the vorticity is $\zeta(x, y) = \frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}$.

The best results were obtained when the cost functional J was chosen to be of the **flow tracking**-type, namely:

$$\mathcal{J}(\Lambda) = \frac{1}{2}\int_{t_1}^{t_2}\int_{\mathbf{D}}(|u-u_d|^2+|v-v_d|^2)d\mathbf{D}dt \tag{7.9}$$

where $\mathbf{D}$ is the spatial domain and (u_d, v_d) is the desired velocity field.

7.4 Existence of the optimal solution

The control problem involving Navier-Stokes equations was studied by Abergel and Temam [1], Coron [39], Fursikov et al. [59].

Ou [154] proved an existence theorem for the optimal controls in the case of a rotating cylinder, continuing the research of Sritharan [178].

First, one needs to construct two solenoidal vector fields $\mathbf{\Psi}(\mathbf{r})$ and $\mathbf{\Phi}(\mathbf{r})$ which would carry the nonhomogenous boundary conditions at the solid surface of the cylinder and, in the far field, respectively. If $\mathbf{r}$ is the position vector and $\mathbf{U}(\mathbf{r},t)$ is the velocity vector for the model equations we introduce a change of variable

$$\mathbf{U}(\mathbf{r},t) = \mathbf{V}(\mathbf{r},t) + U_\infty\mathbf{\Psi}(\mathbf{r}) + \Omega(t)\mathbf{\Phi}(\mathbf{r}) \tag{7.10}$$

where U_∞ is the far field velocity in the x direction and Ω is the angular velocity of the cylinder.

The following system of equations with homogeneous boundary conditios is then obtained:

$$\begin{aligned}\mathbf{V}_t + (\mathbf{V}\cdot\nabla)\mathbf{V} + U_\infty(\mathbf{V}\cdot\nabla\mathbf{\Phi}) + \Omega(t)(\mathbf{V}\cdot\nabla\mathbf{\Psi}) \quad &+ \quad U_\infty(\mathbf{\Phi}\cdot\nabla\mathbf{V}) + \\ \Omega(t)(\mathbf{\Psi}\cdot\nabla\mathbf{V}) = -\nabla P + \frac{1}{Re}\nabla^2\mathbf{V} \ \text{ in } \ \mathbf{D}\times[0,T]& \\ \nabla\cdot\mathbf{V} = 0 \ \text{ in } \ \mathbf{D}\times[0,T]& \\ \mathbf{V}\Big|_{\Gamma} = 0& \\ \mathbf{V}\to 0 \ \text{ as } \ |\mathbf{r}|\to\infty& \\ \mathbf{V}(\mathbf{r},0) = 0&\end{aligned} \tag{7.11}$$

where $\mathbf{D}$ is the domain considered and Γ its boundary.

Let $\mathcal{H}$ be the solenoidal subspace defined by

$$\mathcal{H} = \{\mathbf{V}:\mathbf{D}\to R^2; \mathbf{V}\in L^2(\mathbf{D}), \nabla\cdot\mathbf{V} = 0 \ \text{ and } \ \mathbf{V}\cdot\mathbf{n}\Big|_{\Gamma} = 0\} \tag{7.12}$$

The system of equations (7.11) is projected onto the solenoidal subspace $\mathcal{H}$ by an orthogonal projector and we obtain:

$$\begin{aligned}\delta_t\mathbf{V}(t;\Omega) + \frac{1}{Re}\mathbf{A}\ \mathbf{V}(t;\Omega) + \mathbf{N}(\mathbf{\Psi},\mathbf{\Phi},\mathbf{V}(t;\Omega)) = F(\mathbf{\Psi},\mathbf{\Phi},\Omega)& \\ \mathbf{V}(0) = 0&\end{aligned} \tag{7.13}$$

where $\mathbf{A}$ is the Stokes operator.

We denote $\mathcal{U}_{ad}$ the set of all admissible pair $(\mathbf{V}, \Omega) \in L^2(0,T;\mathcal{H}) \times H^1(0,T)$ that satisfy equation (7.13).

If $\mathbf{z}_d$ is the desired flow field (in our case a flow without Karman vortices) then the optimal control problem is to find an optimal pair $(\mathbf{V}, \Omega) \in \mathcal{U}_{ad}$ which minimizes the cost functional

$$\boldsymbol{\mathcal{J}}(\mathbf{V}, \Omega) = \int_0^T ||\mathbf{V}(t;\Omega) + U_\infty \boldsymbol{\Psi} + \Omega(t)\boldsymbol{\Phi} - \mathbf{z}_d||^2_{L^2(\mathbf{D})}\, dt + \lambda \int_0^T |\Omega_t|^2\, dt \tag{7.14}$$

where λ is a regularization parameter.

The following result was proved by Ou [154] following Sritharan [178]):
There exists an optimal solution $(\mathbf{V}^*, \Omega^*) \in \mathcal{U}_{ad}$ *such that*

$$\mathcal{J}(\mathbf{V}^*, \Omega^*) = \inf_{(\mathbf{V},\Omega)\in\mathcal{U}_{ad}} \mathcal{J}(\mathbf{V}, \Omega) \tag{7.15}$$

7.5 Regularization

Preliminary numerical experiments proved that the minimization is ill-posed (while the objective functional decreased by a very small percentage, the difference in the values of the parameter for which we have this decrease in the function may assume *arbitrarily large values*).

Our approach for dealing with ill-posedness was to apply a Tikhonov-type regularization. We added a new term to the cost functional F:

$$\boldsymbol{\mathcal{J}}_{REG} = \boldsymbol{\mathcal{J}} + \lambda\Pi \tag{7.16}$$

where $\lambda > 0$ is a regularization parameter and Π a regularization function (see Tikhonov and Arsenin [187]).

The regularization term may also be viewed as playing the role of a penalty term aiming to ensure that the control parameter lies within a reasonable interval.

For the case of constant rotation the regularization function Π is:

$$\Pi = \frac{1}{2}\int_\Gamma (u^2 + v^2) \mathrm{d}\Gamma$$

where (u, v) are the two components of velocity and Γ is the boundary of the cylinder.

Such a choice was also made by Abergel and Temam [1] and Gunzburger and Manservisi [90] in their research.

For the time-harmonic case, the regularization function Π was chosen to be:

$$\Pi = \int_0^{T_w} \frac{1}{2}\int_\Gamma (u^2 + v^2) \mathrm{d}\Gamma dt$$

where T_w is the length of the time window for optimization.

An in-depth discussion about regularization is provided in chapter 5.

7.6 Overview of numerical results

The optimization was performed over a short time interval (time window). The values of the state variables for each time step in this control window were saved and used in the adjoint computation (specifically the "forcing term" for the adjoint equation).

The time window was located at the beginning of the time evolution and had a length varying between 1.0 and 4.0 time units.

Even when the flow is considered over a time period of 25.0 time units (which exceeds by far the length of the control time window), the optimized values of the control parameters suppress the Karman vortex shedding far beyond the extent of the time window.

The choice of the length of the time window is very important. For both cases, namely constant and time-dependent angular rotation, the length of the control window should be larger than the vortex shedding period ($\mathcal{VSP}$), the inverse of the Strouhal number $St = \dfrac{f_K D}{U_0}$, where f_K is the Karman vortex street frequency and D is the diameter of the cylinder.

Since the adjoint method requires availability of the values of the state variables for all the time steps in the control time window, the length of the time window should not to be much larger than $\mathcal{VSP}$. Otherwise both the memory and the CPU time requirements for minimization may prove to be too large.

For the case of the constant rotation we obtained satisfactory results with a control time window smaller than $\mathcal{VSP}$ (but not smaller than 1.0 time unit). In the time-dependent case the choice of a time window smaller than $\mathcal{VSP}$ leads to nonconvergence of the minimization process.

The cost functional which was minimized involved the L_2 norm of the difference between the computed velocity and a "desired" velocity. Our "desired" flow was obtained for Reynolds number $Re = 2$ and the ratio between the angular velocity and the free stream velocity had a value of 2.0 (see Figure 7.6).

7.7 Suppression of Karman vortex shedding in the constant rotation case

Let us consider the speed ratio

$$\alpha = \frac{a\Omega}{U},$$

where a is the radius of the cylinder, Ω is the angular velocity and U is the free stream velocity.

The uncontrolled flow is taken at $\alpha = 0.5$ (an example is provided in Fig. 7.7, for $Re = 100$). The minimization satisfies the convergence criteria after 5-11 minimization iterations for all the cases we considered: the Reynolds number in the range $60 \leq Re \leq 1000$.

For each case considered we found a threshold value for α (denoted α_{Re}) such that for any $\alpha > \alpha_{Re}$ a full suppression of the Karman vortex shedding was obtained (see Figures 7.8,7.9, 7.10).

The CPU time required for a typical optimal flow control calculation was 2-3 hours on a Silicon Graphics Indigo (SGI) machine.

The results for $60 \leq Re \leq 160$ were found to be in very good agreement with the numerical results obtained by Kang et al. [118] (see Fig. 7.3).

For the case $60 \leq Re \leq 140$ the regularization parameter was found by using an empirically derived law, which relates it to the Reynolds number (see Fig. 7.5). We started by finding the values of the regularization parameter by trial and error for two Reynolds numbers (we considered $Re = 60$ and $Re = 100$) and then we assumed the existence of a logarithmic relation between the regularization parameter and the Reynolds number. Based on this assumption we were able to obtain the corresponding regularization parameters for the other Reynolds numbers (in our case $Re = 80$, $Re = 120$ and $Re = 140$, respectively).

For the case $160 \leq Re \leq 1000$ the empirical law employed in the previous case for obtaining the regularization parameter did not yield good results and, as a consequence, the corresponding regularization parameters were found by trial and error. A possible explanation of this phenomenon is the following: the Karman vortex regime for $160 \leq Re \leq 1000$ is inherently different than the regime for $60 \leq Re \leq 140$ (Zdravkovich [209]).

To check that the minimization results were robust, we performed for each case two different minimizations: one starting with an initial guess of $\alpha = 0.9$ (a value less than the optimal value) and one starting with an initial guess of $\alpha = 3.5$ (a value greater than the optimal value of α). For both initial guesses, the results obtained for the optimal value of α were identical.

As the Reynolds number increases from 60 to 1000 we can see from Figure 7.4 that the rotation rate tends asymptotically to a limit. This behaviour is in good agreement with previously obtained experimental and numerical results.

At $Re = 1000$ we compared our results with the values obtained by Chew et al. [29]. They found that for $\alpha = 2$ and $\alpha = 3$ any vortex shed will be weak and Karman vortex shedding almost disappears for $\alpha = 3$, a phenomenon which was also described experimentally by Badr et al. [9] and numerically by Chou [32]. We found the "optimal" α to be $\alpha = 2.32$ for $Re = 1000$.

For $Re \geq 200$ the flow is not completely free of vortex shedding (as it can be seen from Fig. 7.9 and 7.10). This situation was also described by Chen et al. [28].

In the case presented here (time independent angular velocity) we found that control time windows smaller than the Karman vortex shedding period (but not smaller than 1.0 time units) gave satisfactory results. This observation is important since a smaller control window reduces the computer memory necessary for storing the state variables (which are required for the adjoint computation). A smaller time window also means a sizable reduction in the required CPU time.

7.8 The time histories of the drag coefficient in the constant rotation case

Practical applications (in aerodynamics) of optimal control for flow around a rotating cylinder involve the optimization of the drag coefficient (C_D).

We compare the variation of the drag coefficient in the controlled case (with rotation) with the corresponding variation for the no-rotation case ($\alpha = 0$). In order to compare them on the same plot we subtracted from C_D the corresponding mean value ($\bar{C}_D$). The mean drag coefficients obtained numerically for the case of no rotation were in agreement with the values reported by He et al. [98] (see Table 7.1).

We noticed a very significant reduction in the amplitude of the fluctuation for the drag coefficient when the flow is controlled.

In a viscous flow the total drag forces are contributed by the pressure and skin friction due to the viscous effects. For known vorticity values

$$\omega(x, y) = \frac{\partial u}{\partial y} - \frac{\partial v}{\partial x}$$

on the cylinder surface, the drag can be calculated in the polar coordinates $r - \theta$:

$$C_D(t) = C_{D_P}(t) + C_{D_f}(t) = \frac{2}{Re} \int_0^{2\pi} \Big[(\frac{\partial \omega(t)}{\partial r} \Big]_\Gamma \sin\theta d\theta - \Big[\frac{2}{Re} \int_0^{2\pi} \omega(t) \Big]_\Gamma \sin\theta d\theta \quad (7.17)$$

where the subscript Γ denotes quantities evaluated on the cylinder surface and the subscripts P and f represent the contributions from pressure and friction, respectively.

Fig. 7.14 shows plots of the time histories of the drag coefficient for different Reynolds numbers and for time in the interval $0 \leq t \leq 20$ time units. On each plot we present 2 graphs: the drag obtained for a flow in the fixed cylinder case ($\alpha = 0$) and, respectively, the drag for the flow obtained using the optimal value of the control speed ratio α (after subtracting the corresponding mean value).

The results presented demonstrate the effectiveness in improving the drag performance by selecting a proper rotation rate. An example is presented in Fig. 7.14, which shows a reduction of more than 60% of the amplitude of the drag variation.

7.9 Suppression of Karman vortex shedding for the time harmonic rotary oscillation

We also considered the time dependent angular velocity. A special case is the time harmonic rotary oscillation, for which the speed ratio assumes the form $\alpha(t) = A \sin(2\pi F t)$.

The minimization was performed for values of the Reynolds numbers in the range $100 \leq Re \leq 1000$.

Several time control windows were used (their length of the control varying between 1.0 and 5.0 time units). In order to obtain numerical convergence for the minimization we had to choose a time window longer than the Karman vortex shedding period, otherwise the minimization failed to converge.

The regularization parameter was chosen by trial and error. For this case we could not find a relationship between the regularization parameter and the Reynolds number, as for the previously described constant rotation rate case.

The flow obtained using the optimal values of the angular velocity after the minimization is presented in Fig. 7.12 and 7.13. In this case we did not obtain complete suppression of the vortex shedding.

However, we can see that the flow determined by the optimal rotations parameters (obtained through the minimization process) is markedly less turbulent than the uncontrolled flow, described in Fig. 7.11.

7.10 The time histories of the drag coefficient for the time harmonic rotary oscillation

Reduction of the drag coefficient using time harmonic rotary oscillation was reported by Tokumaru and Dimotakis [188], Baek and Sung [10] and He et al. [98]. The research of He et al. [98] shows a 30% to 60% drag reduction if one uses a rotating cylinder, compared to the fixed cylinder configuration.

Our results are presented in Fig. 7.15 which show plots of the time histories of the drag coefficient for time in the interval $0 \leq t \leq 20$ time units.

They are not as impressive as the results obtained for the constant rotation case, which may be related to the fact that we could not obtain the full suppression of Karman vortex shedding.

If one compares our results with He et al. [98], one may distinguish small differences in the numerical values obtained for the optimal control parameters (in both research articles, the forcing angular velocity is $\omega(t) = \omega_1 \sin(2\pi S_e t)$ and the optimal control parameters are the amplitude ω_1 and the forcing frequency S_e). Our "optimal" amplitude ω_1 differs by at most 10% from the value reported in their research. We did not obtain the same "optimal" forcing frequency (which in their case was very close to the lock-in forcing frequency).

One possible explanation for this situation is the following: there is a difference in the formulation of the cost functionals used in our research and those described in He et al. [98] (this difference appears to be due to the setting of the optimal control problem: our main goal was the suppression of the Karman vortex shedding, while their research was aimed toward reduction of drag).

7.11 Description of the physical phenomena corresponding to the computational results

At low Reynolds numbers ($Re < 40$) the wake behind a non rotating cylinder comprises a steady recirculation region with two vortices symmetrically attached to the cylinder, whose size grows with increasing Reynolds number. When the Reynolds number is slightly larger, $Re < 60$, the trailing vortex street becomes unstable and develops an unsteady wavy pattern. For Reynolds numbers $60 < Re < 200$, the Karman vortex shedding occurs in the near wake behind a cylinder due to the flow instability accompanying a large fluctuating pressure and, thus, a periodically oscillating lift force. The attached vortices become asymmetric and are shed alternately at a well defined frequency.

At higher Reynolds numbers ($Re > 200$) the flow becomes more turbulent and vortex shedding also occurs, but assuming more complicated patterns this time. In this last case the vortex structures are unstable to 3-D perturbations. For this reason, numerical results available from the 2-D codes agree well with the experimental data for Reynolds numbers $Re \leq 160$ bwhile numerical results obtained for larger Reynolds numbers are not always consistent as a consequence of the three-dimensionality effect (e.g., Graham [76]).

For higher Reynolds numbers 3-D codes will yield numerical results which will match experimental data better than their 2-D counterparts. Zhang and Dalton [210] obtained smaller global quantities such as drag and lift (with better agreement with experimental values) than the corresponding 2-D simulation. The difference has been attributed to the phase difference of flows in different spanwise locations caused by three-dimensionality and to the 3-D mixing, both absent in the 2-D simulation.

For Reynolds numbers $Re \geq 160$ there are various instabilities. The primary instability can be seen when the wake undergoes a supercritical Hopf bifurcation that leads to 2-D Karman vortex street. The secondary instability occurs sequentially, which results in the onset of the 3-D flow. The periodic wakes are characterized by two critical modes which are respectively associated with large-scale and fine-scale structures in span (Williamson [205], Ding and Kawahara [48]).

The rotation of a cylinder in a viscous uniform flow is expected to modify the wake flow pattern and vortex shedding configuration, which may reduce the flow-induced oscillation or augment the lift force. The basic physical rationale behind the rotation effect is that as the cylinder rotates, the flow of the upper cylinder is decelerated and easily separated, while the flow of the lower cylinder is accelerated and the separation can be delayed or suppressed. Hence the pressure on the accelerated side becomes smaller than that of the decelerated side, resulting in a mean lift force (this effect is known as "Magnus effect": Barkla and Auchterlonie [11]).

As we increase the control parameter α (the angular velocity normalized by the free stream velocity), the flow becomes asymmetric and at the same time the pressure on the lower (accelerated) side of the cylinder decreases, resulting in a negative downward mean lift. The rotation effect is mainly confined to the flow in the vicinity of the cylinder surface. For the near-surface flow, as α increases, the negative vorticity on the upper side of the cylinder dominates the positive vorticity on the lower side, thus weakening the vortex shedding which eventually disappears.

There is a transition state (called *critical* state) between the state of periodically alternate double side shed vortex pattern for smaller α and the state of steady single side attached vortex pattern for larger α (e.g., Ling and Shih [133], Badr et al. [9], Chen et al. [28]).

Another characteristic of the flow is the synchronization between cylinder and wake. This will determine the apparition of a "lock-on" phenomenon. In the case of time harmonic rotary oscillations this phenomenom was described experimentally by Tokumaru and Dimotakis [188] and numerically by Chou [31] and Dennis et al. [46] (who studied the effects of the forcing frequency and amplitude on a cylinder wake).

The combined system of cylinder and wake will be locked in if the forcing frequency lies in the neighborhood of the natural Karman frequency. According to He et al. [98], the natural Karman frequency is the optimal value for the forcing frequency for the drag reduction.

For this case (time dependent rotational oscillation) two co-rotating vortex pairs are shed away from the cylinder to form a co-rotating vortex pair which slows down their convection further downstream. This seems to delay the development of the periodic flow pattern in the near wake.

We have two phenomena when the forcing frequency is lower than the natural shedding frequency. An initial clockwise vortex is formed on the lower half of the cylinder when the cylinder is rotated in the counterclockwise direction while a counterclockwise vortex is formed on the upper half when the clockwise rotation starts. A non-synchronized vortex formation mode is developed which cannot lead to suppression of Karman vortex shedding.

One can also distinguis two vortices when the forcing frequency is higher than the natural shedding frequency. An initial reactive clockwise vortex is formed on the upper half of the cylinder when the cylinder is rotated in the counterclockwise direction while counterclockwise vortex is formed on the lower half when the clockwise rotation starts. This leads to a synchronized vortex mode, which is one of the reasons why the optimal values for the forcing frequency obtained in the previous section cannot be lower than the vortex shedding frequency.

The behavior of the drag coefficient C_D is determined by the fact that flow separation is a major source of pressure drag and the moving-wall effects will postpone this separation. As shown by Prandtl in 1925 [161] separation is completely eliminated on the side of the cylinder where the wall and the freestream move in the same direction while on the other side of the cylinder separation is developed only incompletely.

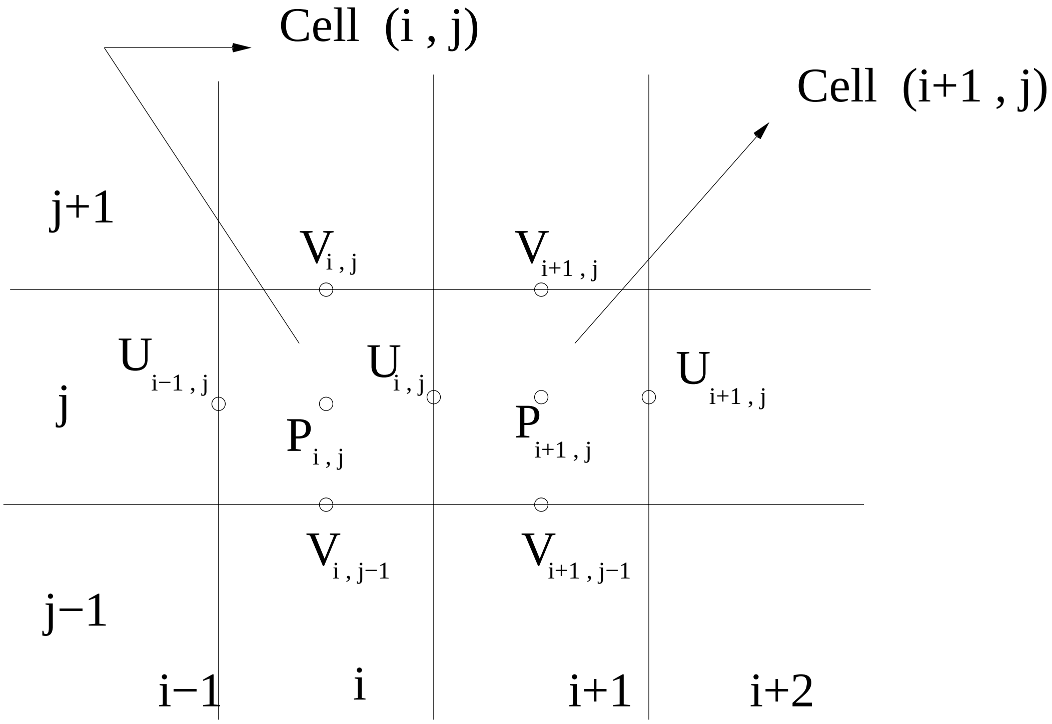


Figure 7.1. Staggered grid

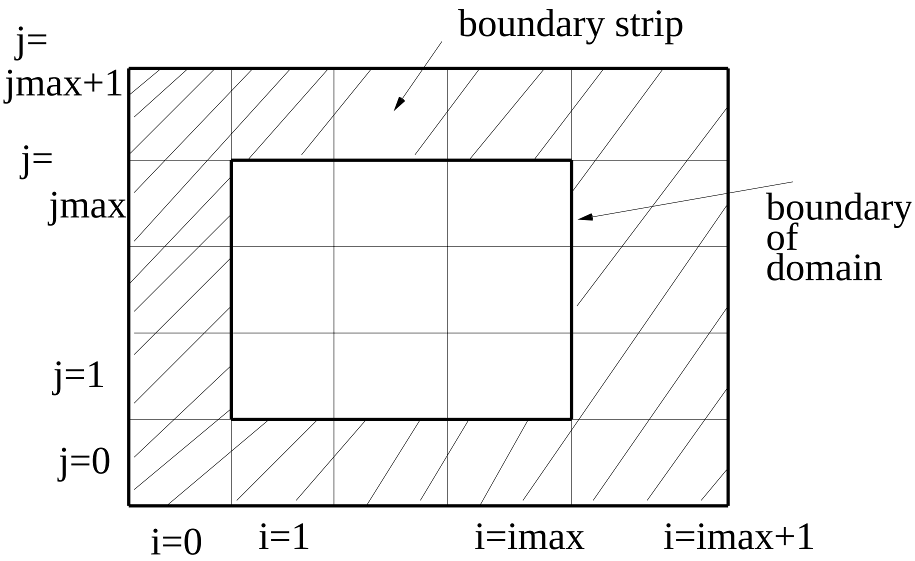


Figure 7.2. Domain with boundary cells

Table 7.1. The mean value of the drag coefficient $\bar{C}_D$ for various Reynolds numbers

Re	100	200	400	700	1000
Present work	1.42	1.44	1.54	1.59	1.68
He et al. [98]	1.35	1.36	1.42	1.48	1.52

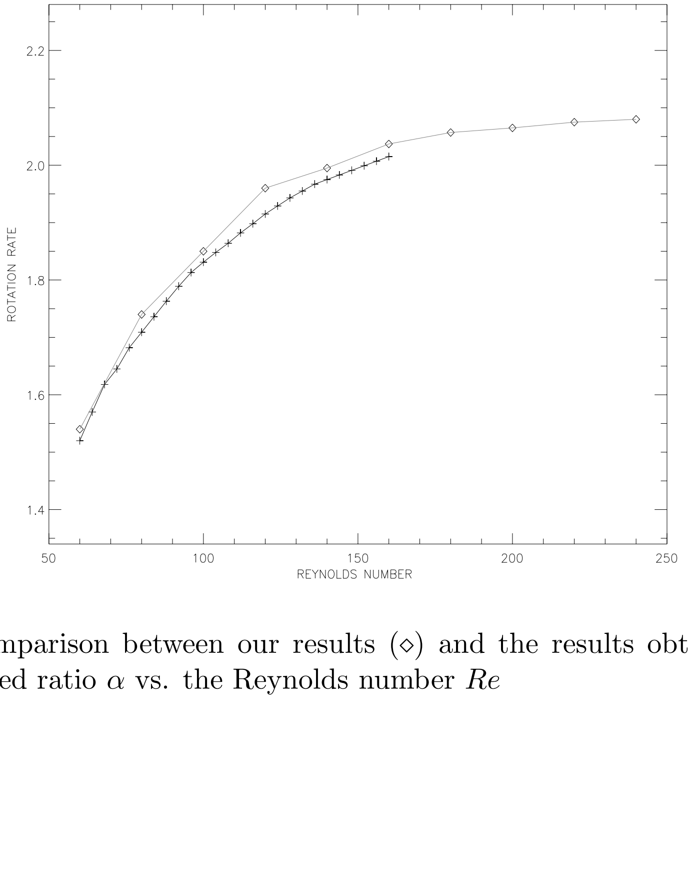


Figure 7.3. Comparison between our results (◇) and the results obtained by Kang et al.(1999): the speed ratio α vs. the Reynolds number Re

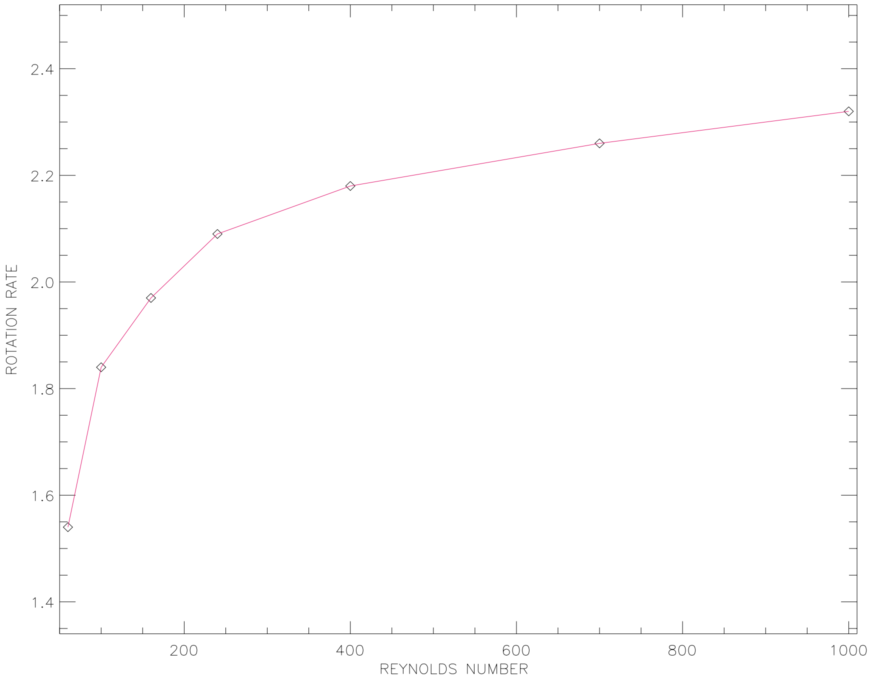


Figure 7.4. The optimal speed ratio α vs. the Reynolds number *Re*

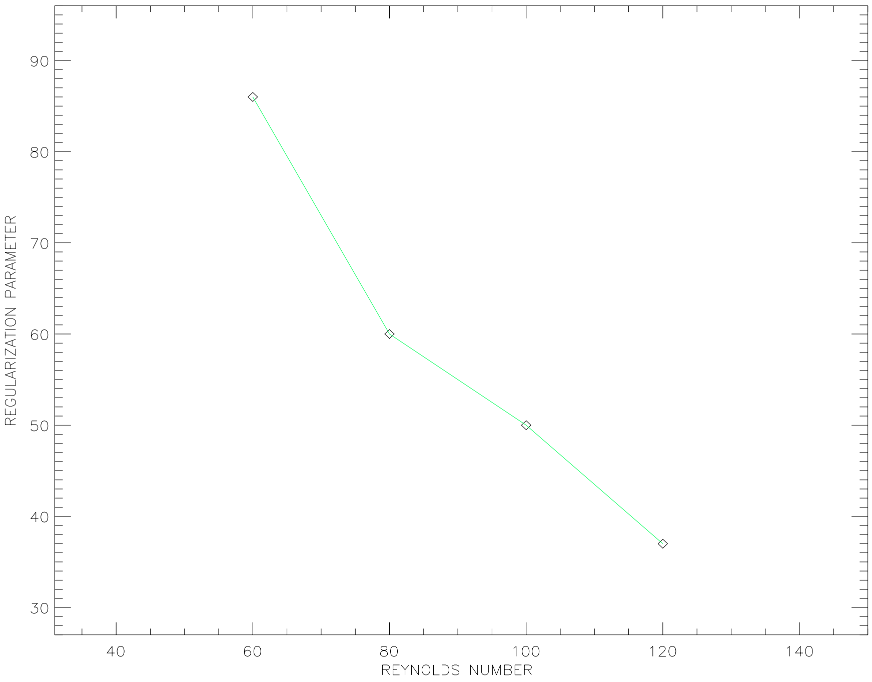


Figure 7.5. Regularization parameter vs. Reynolds number *Re*

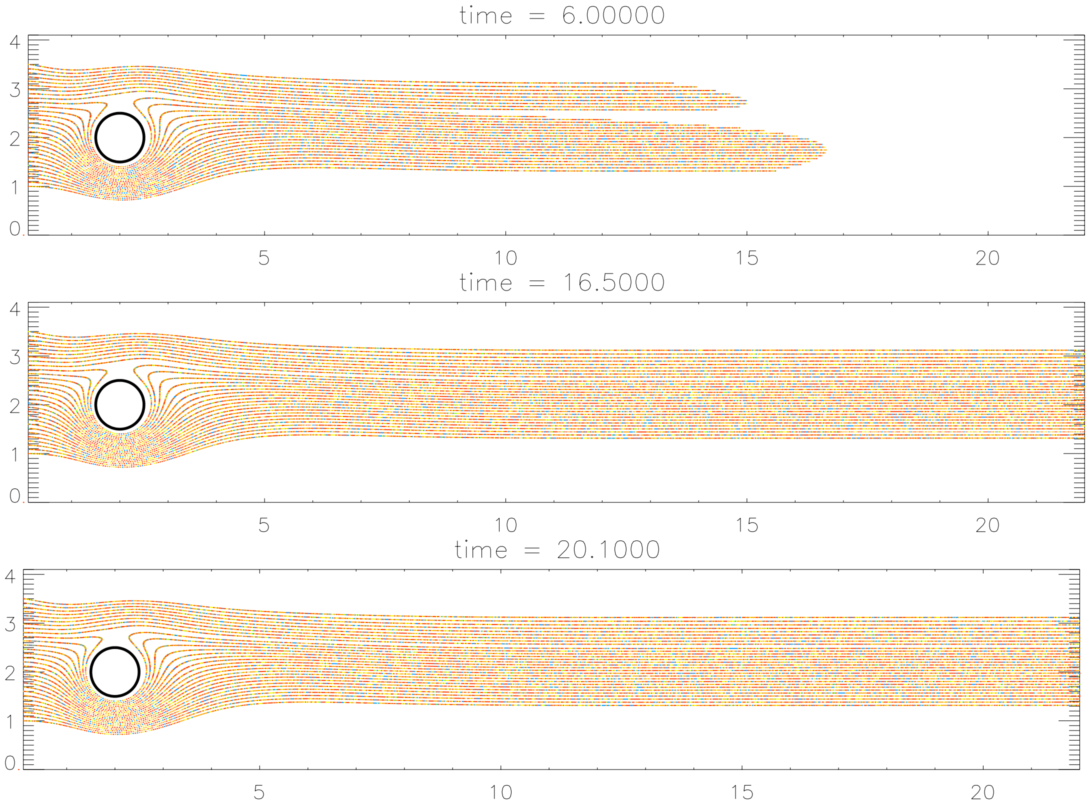


Figure 7.6. Streaklines for the "desired" flow at $Re = 2$ and speed ratio $\alpha = 2.0$

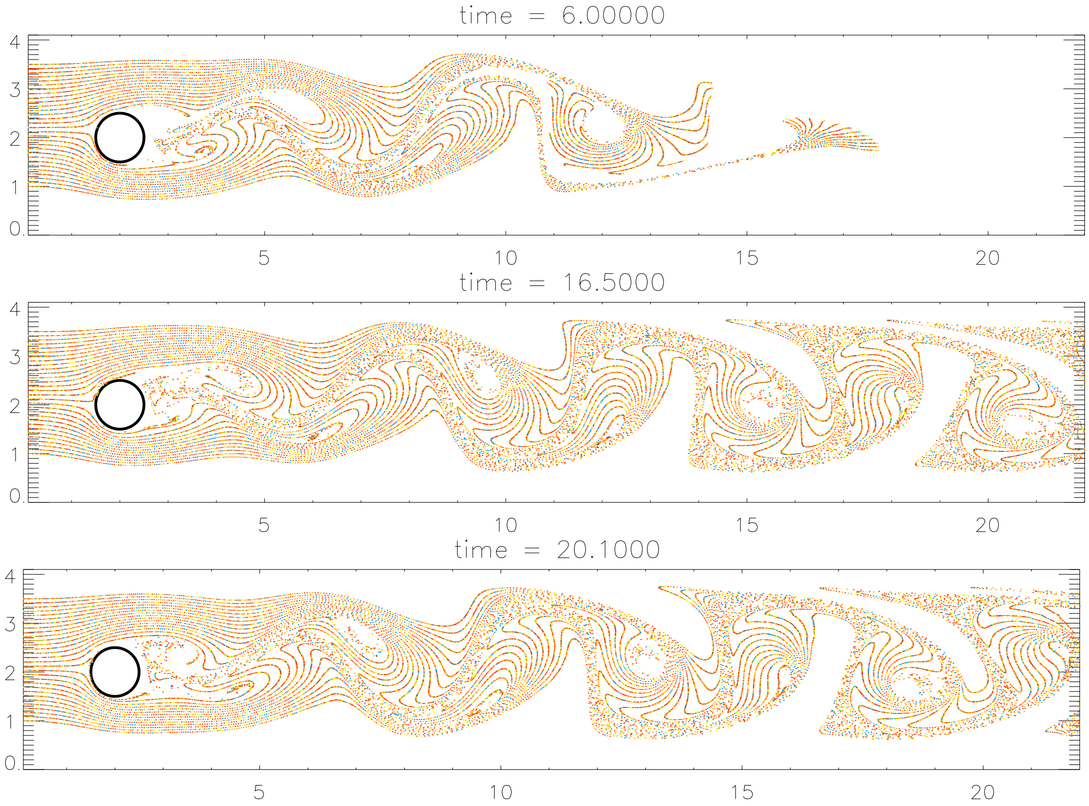


Figure 7.7. Streaklines for uncontrolled flow at $Re = 100$ and speed ratio $\alpha = 0.5$

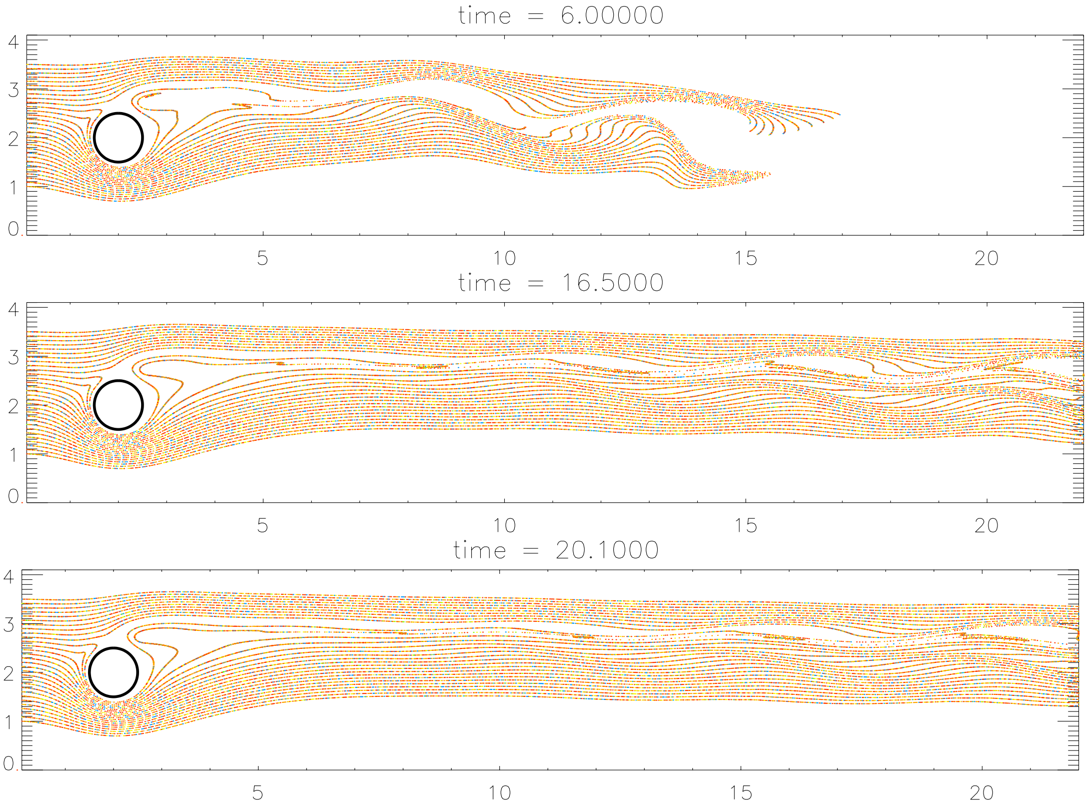


Figure 7.8. Streaklines for controlled flow at $Re = 100$ with optimal speed ratio $\alpha = 1.84$

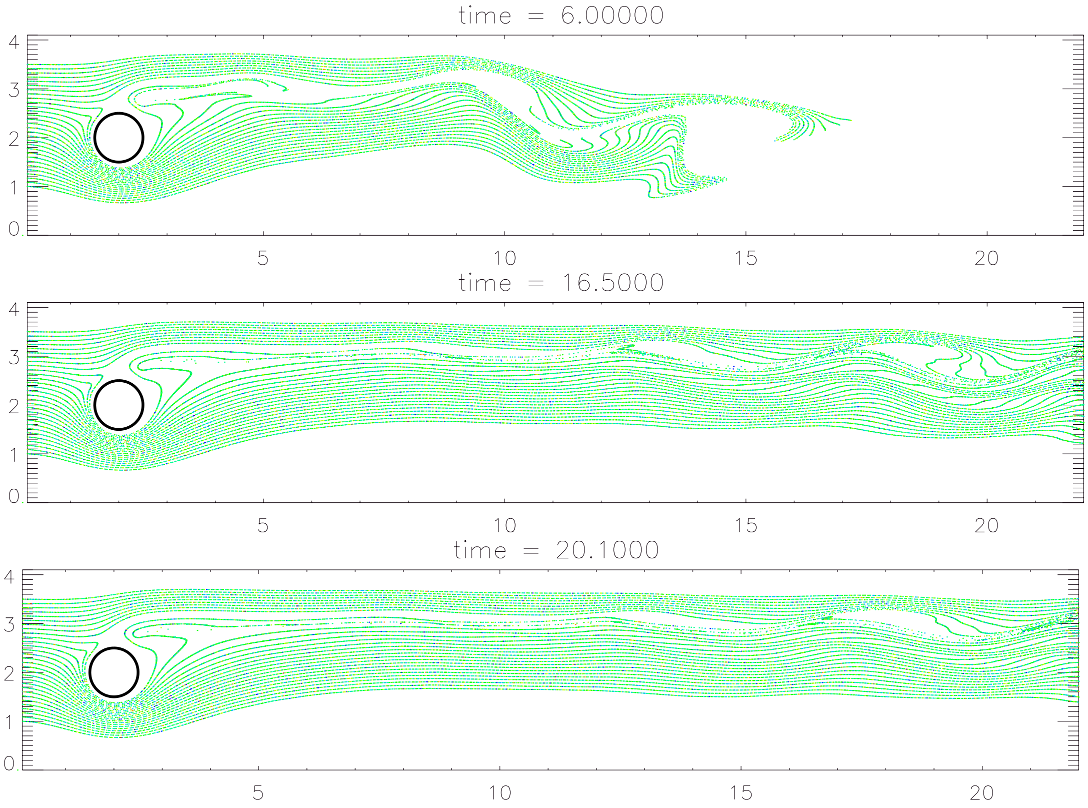


Figure 7.9. Streaklines for controlled flow at $Re = 400$ with optimal speed ratio $\alpha = 2.18$

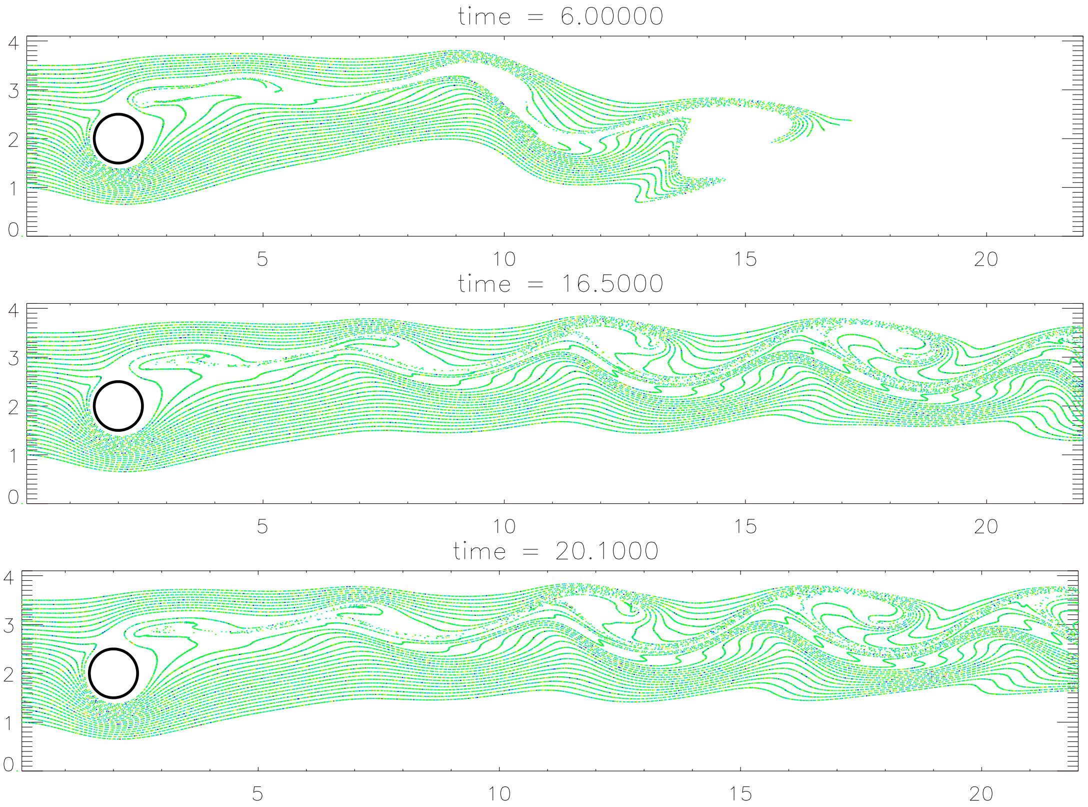
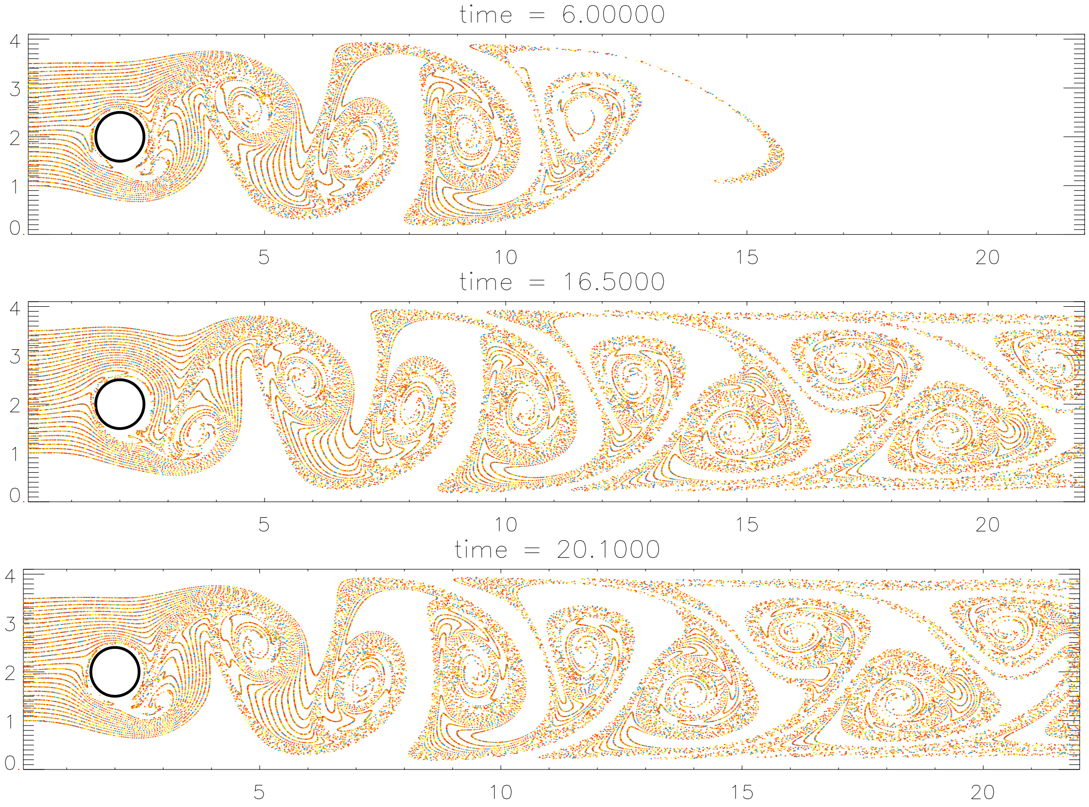


Figure 7.10. Streaklines for controlled flow at $Re = 1000$ with optimal speed ratio $\alpha = 2.35$



time = 6.00000

time = 16.5000

time = 20.1000

Figure 7.11. Streaklines for the uncontrolled flow at $Re = 100$ and speed ratio $\alpha(t) = 2.5 \sin(1.0\pi t)$

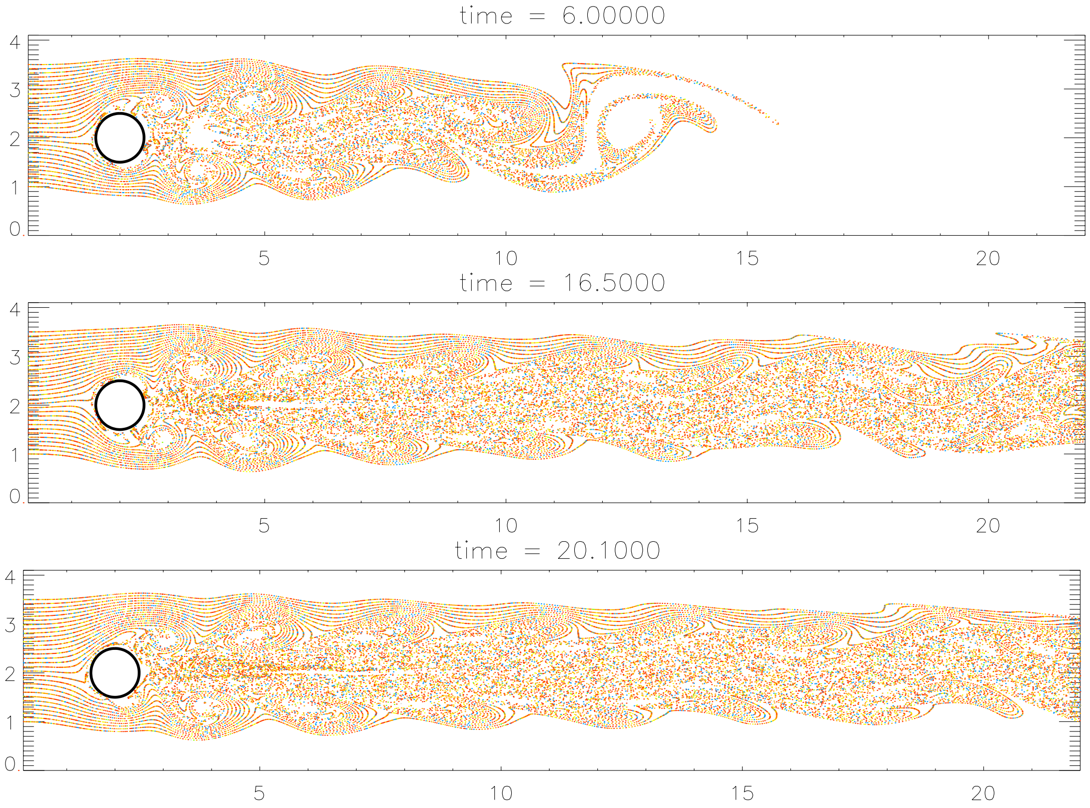


Figure 7.12. Streaklines for the controlled flow at $Re = 100$ with optimal parameters $A = 6.5$ and $F = 1.13$; $\alpha(t) = A \sin(2\pi F t)$

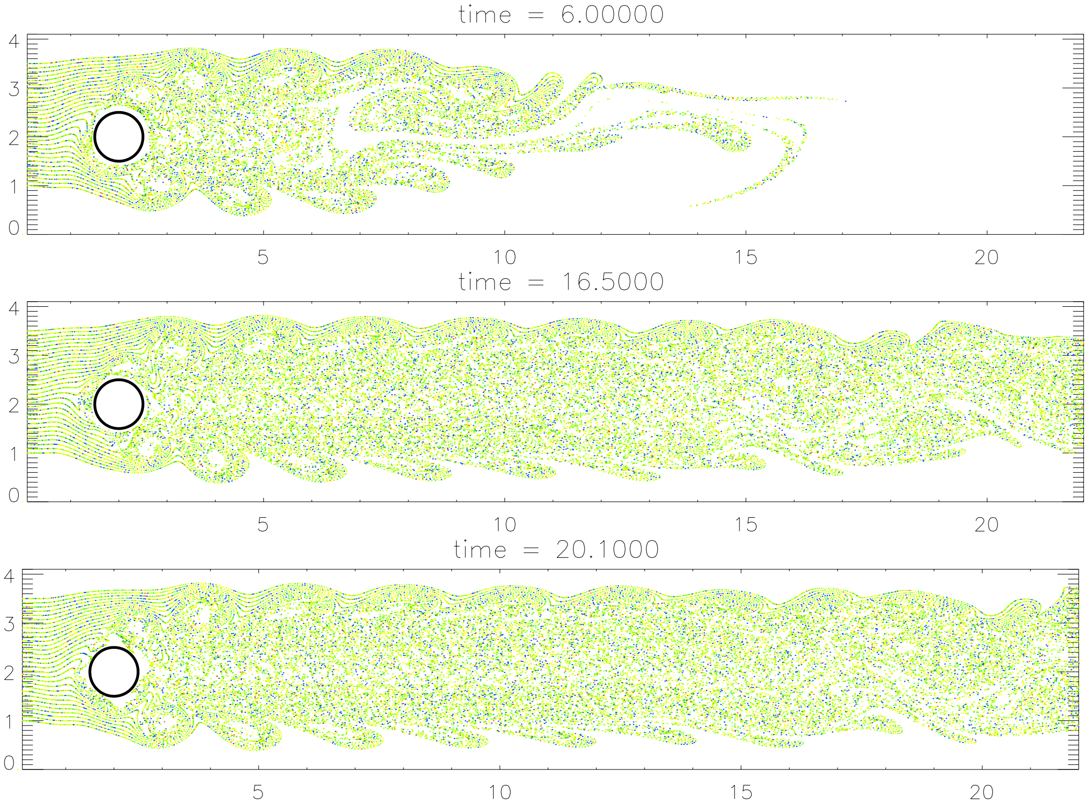
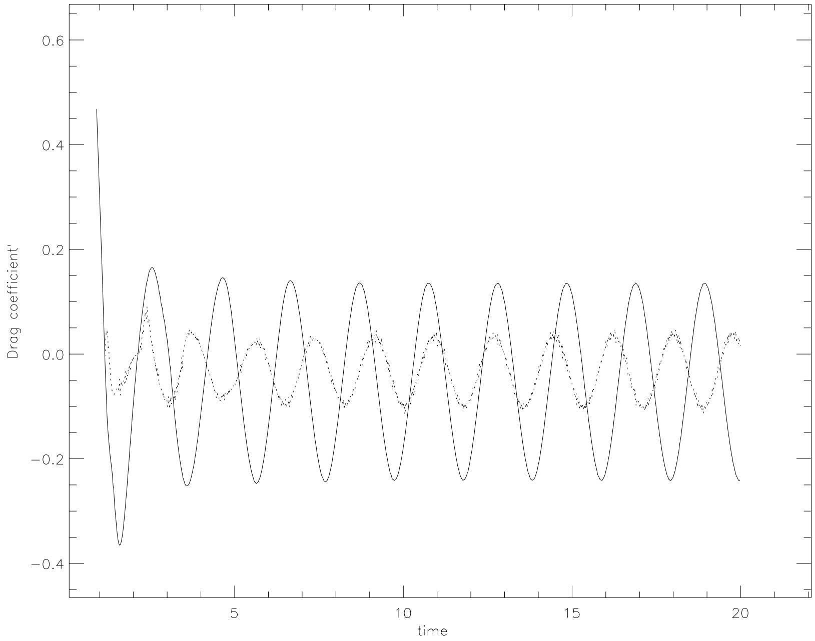
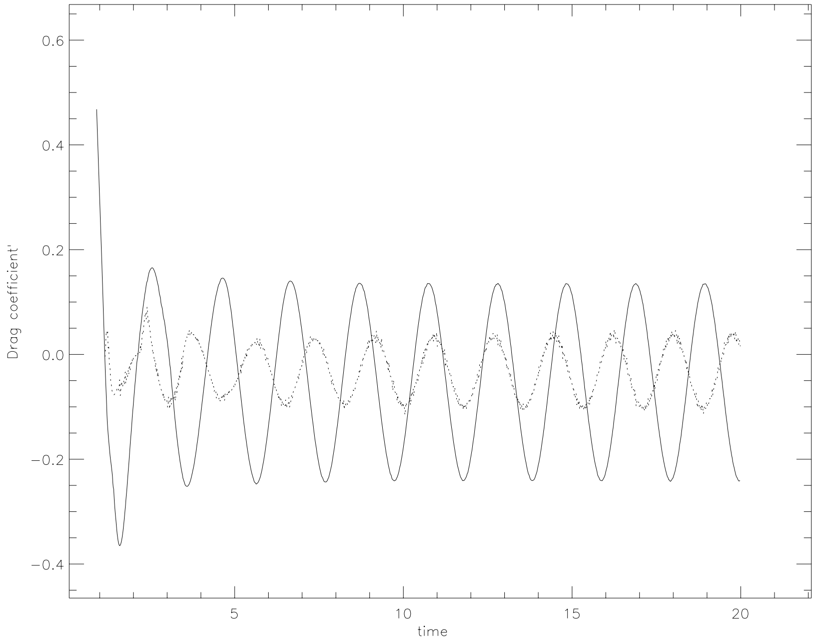


Figure 7.13. Streaklines for the controlled flow at $Re = 1000$ with optimal parameters $A = 6.0$ and $F = 0.86$; $\alpha(t) = A \sin(2\pi F t)$

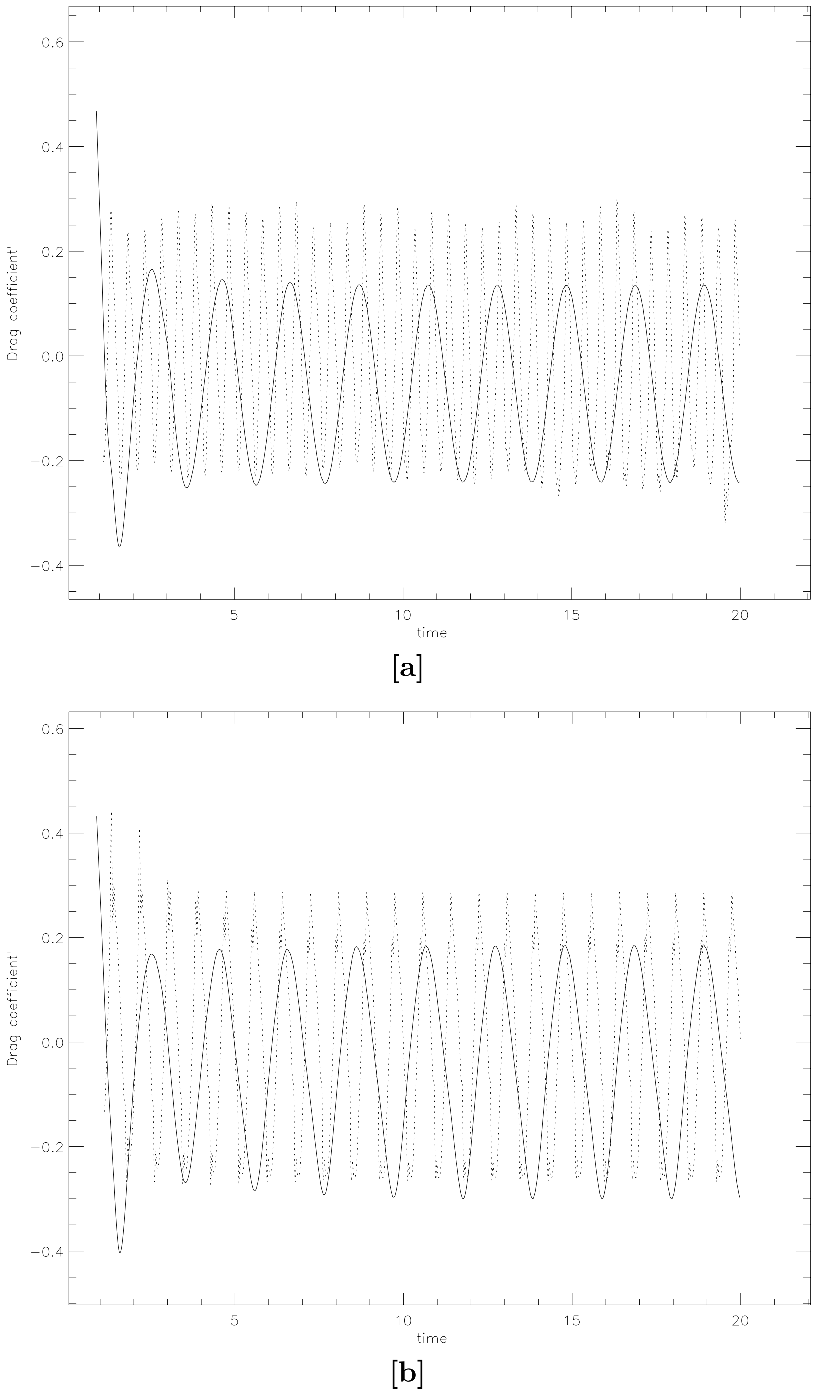


[a]



[b]

Figure 7.14. The variation of the drag for the constant rotation in the controlled (dotted line) and uncontrolled case (continuous line) for **[a]** $Re = 100$ and **[b]** $Re = 1000$



[a]

[b]

Figure 7.15. The variation of the drag for the time-dependent speed ratio $\alpha(t)$ in the controlled (dotted line) and uncontrolled case (continuous line) for **[a]** $Re = 100$ and **[b]** $Re = 1000$

CHAPTER 8

DESCRIPTION OF THE PHYSICAL PHENOMENA FOR THE SHOCK-TUBE PROBLEM

The shock-tube example corresponds to the 1-D Riemann problem for the Euler equations. Its mathematical formulation will be discussed in more detail in chapter 10. This problem was chosen since it contains many "troublesome" aspects present in typical flow solutions, including shock waves, rarefaction waves and contact discontinuities.

The shock tube is also extensively used in studying unsteady short-duration phenomena in varied fields of aerodynamics, physics and chemistry. The transient wave phenomena when a shock wave propagates at a high speed, as well as wave structure and wave interactions, can be studied in shock tubes. Because of high stagnation enthalpies (and temperatures) that are attained, the shock tube provides means to study the thermodynamic properties of gases at high temperatures, dissociation, ionization and chemical kinetics.

The shock-tube consists of a long duct of constant cross-section divided into two chambers by a diaphragm, as shown in Fig. 8.1. The left chamber, called the *driver section*, contains gas at high pressure whereas the right chamber, called the *expansion section*, contains gas at a low pressure. The low-pressure gas may be the same as or different from the high-pressure gas.

At time $t = 0$ the diaphragm is ruptured and a series of compression waves rapidly coalesces into a normal shock wave. The pressure distribution at $t = 0$ is a "step" function. The variables are denoted by V (velocity), ρ (density), P (pressure) and T (temperature). The wave propagates at supersonic speed in the expansion chamber and sets up the gas behind it in motion in the direction of the shock at velocity V_2 (the subscripts correspond to the regions of the flow shown in Fig. 8.3). The laws of normal shock dictate that $P_2 > P_1, T_2 > T_1$ and $\rho_2 > \rho_1$. At the same time a rarefaction wave emanates at the diaphragm section and propagates in the opposite direction into the driver section (4). The leading rarefaction wave (head wave) propagates into the gas of the driver section at a local speed of sound of c_4.

Similarly, the tail wave propagates at a local speed of sound of c_3. The gas behind the last rarefaction wave (tail wave) is set in motion to the right at a velocity V_3 equal to V_2. The shock wave and the rarefaction wave interact in such a manner to establish a common pressure $P_2(= P_3)$ and a common velocity $V_2(= V_3)$ for the gas downstream of these waves. The velocity V_2 can be either subsonic or supersonic.

The gases in regions (2) and (3) differ, however, in temperature and entropy. This creates a surface of discontinuity which moves to the right at the same velocity of the gases in these two regions. The distributions of properties of the gas along the shock tube at a later time $T = 0.24$ are shown in Fig. 8.5. As shown in the figure, the velocity and pressure change in a continuous fashion between regions (3) and (4), owing to the passage of the expansion wave. These properties, however, change discontinuously between regions 1 and 2 as a result of the passage of the shock wave. The trajectories of fluid particles on both sides of the diaphragm are represented in Fig. 8.4 by *abc* and respectively $a^{'}b^{'}c^{'}$.

The properties in the regions (2) and (3) need to be determined in terms of the initial properties in chambers (1) and (4). The analytical solution is presented in chapter 10 using a basic parameter of the shock tube: the *diaphragm pressure ratio* P_4/P_1. The properties in region (2), which is at a higher temperature than region (3), remain uniform until the passage of the reflected waves from either side of the tube or until the passage of the contact surface. The Mach number in this region increases as the ratio P_4/P_1 increases. Unlike the flow across the shock wave, flow across the rarefaction wave is isentropic and these waves propagate into region (4) at the local speed of sound.

The region of the fluid which is traversed by the shock has the index (2) while the region traversed by the expansion wave is denoted by (3), as seen in Fig. 8.3. The interface between regions (2) and (3) is called the *contact surface.* It marks the boundary between the gases which were initially on either side of the diaphragm. Neglecting diffusion, they do not mix, but are permanently separated by the contact surface (which is like the front of a piston, driving into the low-pressure chamber).

On either side of the contact surface the temperatures and the densities may be different but it is necessary that the pressure and the velocity to be the same. These conditions are sufficient to determine the *shock strength* P_2/P_1 and the *expansion strength* in terms of the diaphragm pressure ratio P_4/P_1. Once the shock strength is known, all other flow quantities are easily determined from the normal shock relations.

Although the values of the velocity and pressure across the shock and expansion must be identical, this is not necessarily true for the density and temperature, and in fact they are different. The temperature behind the expansion wave is given by the isentropic relation while the temperature behind the shock is given by the Rankine-Hugoniot relation.

Experimentally it is not possible to start the flow the ideal way, since the bursting or shattering of the diaphragm is a complicated, three dimensional phenomenon. Nevertheless a plane shock is developed within few diameters, by the steepening effect associated with compression waves. The duration of flow is limited by the lengths of the expansion and compression chambers, since the shock wave and expansion wave reflect from the end of the chambers and eventually interact with each other.

Many applications of the shock tube have been discovered. For instance, the uniform flow behind the shock may be used as a short-duration wind tunnel. In this role the shock tube is similar to an intermittent (blow-down tunnel), with the difference that the duration of flow is much shorter.

The abrupt changes of flow condition at the shock front have been utilized for studying transient aerodynamic effects and for studies of dynamic and thermal response.

In the field of molecular physics the shock tube model provides a simple tool for producing fast changes in the state of a fluid in order to observe relaxation effects, reaction rates, etc. In addition, dissociation and ionization were studied using the high enthalpies that were obtained in a shock tube.

As mentioned in the introduction, the shock-tube problem was chosen since it contains many characteristics of physical problems with discontinuities. We presented the dynamics of shocks only for the 1-D case. Aan extensive literature exists for the physical phenomena related to shock dynamics in 2-D or 3-D.

For many practical applications where the shocks are found (e.g., aerodynamics Anderson [5]) it is very important to study not only the shocks alone but also their interaction with other organized structures of the flow: vortex or temperature (Erlebacher and Hussaini [53]), axisymmetric entropy or temperature spot (Hussaini and Erlebacher [107]), vortices (Erlebacher et al. [54], [55]).

Based on these interactions one may consider a different optimal control problem than the one considered for our research: for example, one may control the nonlinear effects of the interactions such that only the most desired characteristics of the flow are kept at the end of the process of optimal control.

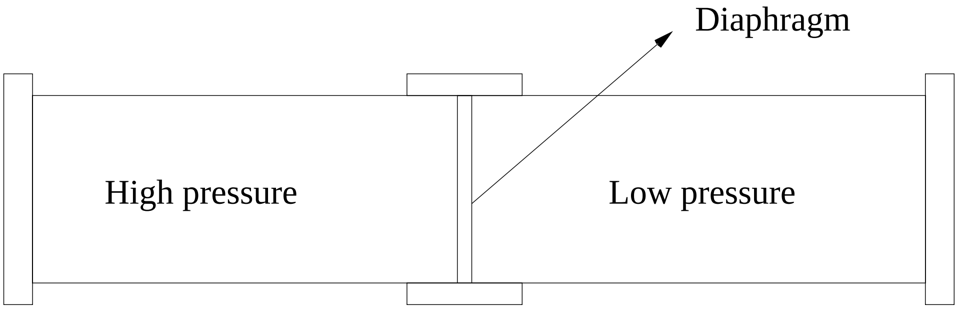


Figure 8.1. The shock-tube problem at time t=0

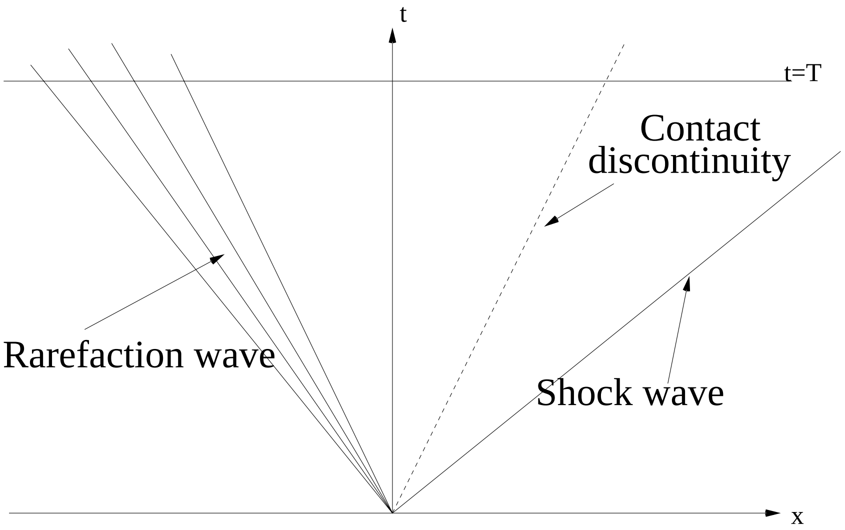


Figure 8.2. The solution shock-tube problem at time $t = T$

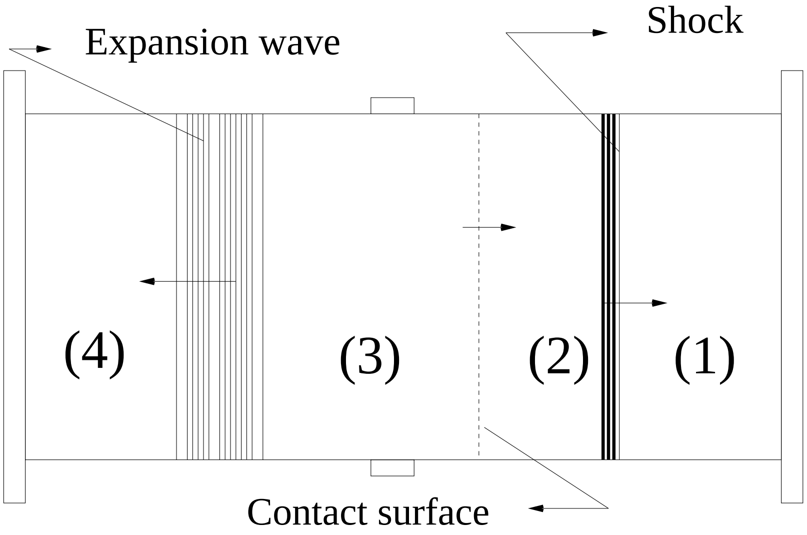


Figure 8.3. Evolution of the flow for the shock-tube problem

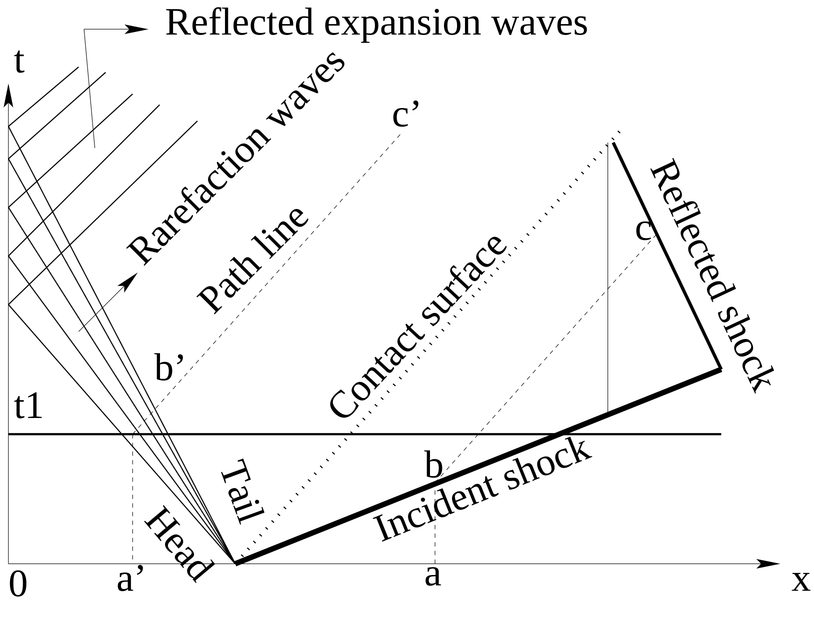


Figure 8.4. The trajectory of the fluid particles for the shock-tube problem

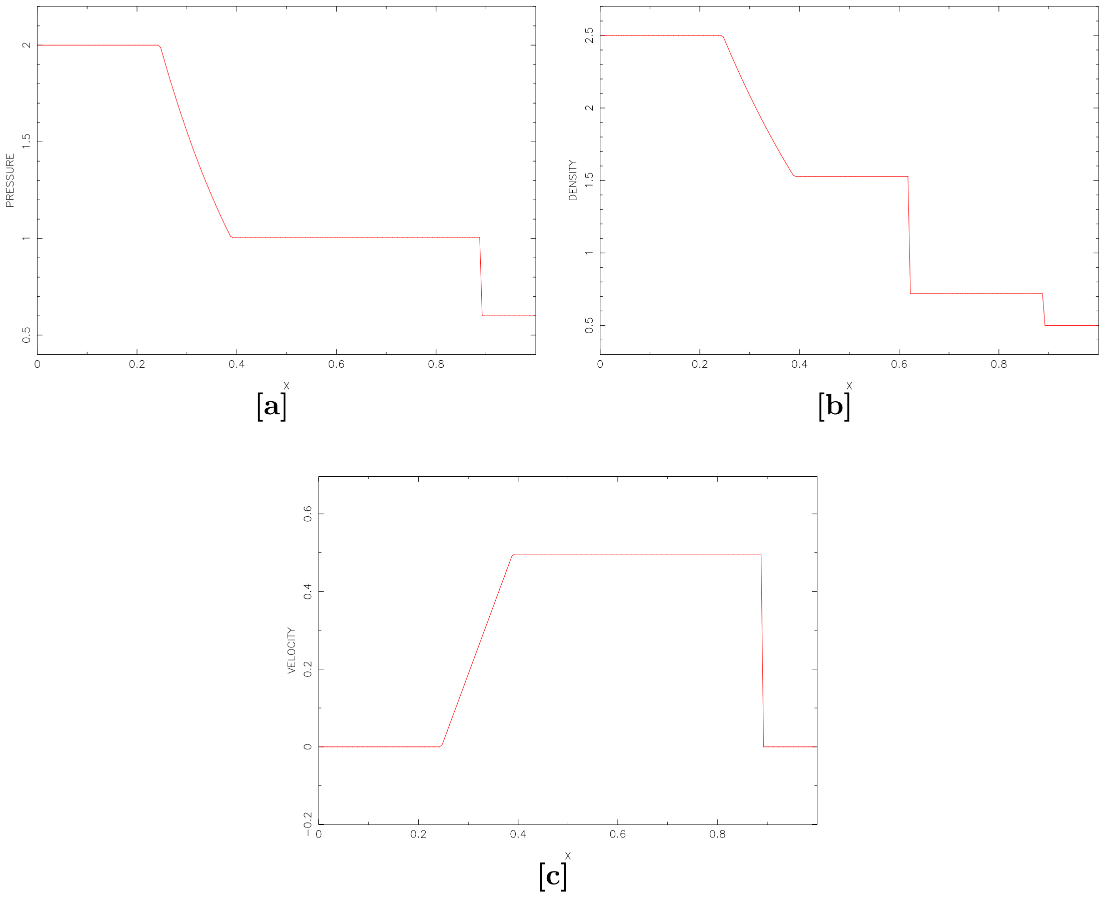


Figure 8.5. Exact solution of the shock-tube problem at time $t = 0.24$: [**a**] pressure, [**b**] density and [**c**] velocity

CHAPTER 9

SENSITIVITIES FOR A FLOW WITH DISCONTINUITIES

9.1 Model formulation

We chose to perform linearization of 1-D Euler equations and sensitivity computation for this discontinuous flow, since the one-dimensional shock-tube problem from gas dynamics contains many potential "troublesome" characteristics of a flow with discontinuities, including shock waves, rarefaction waves and contact discontinuities.

The one-dimensional equations of gas dynamics can be written in conservation law form as:

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = 0 \tag{9.1}$$

where

$$\mathbf{U} = \begin{bmatrix} \rho \\ m \\ e \end{bmatrix}, \qquad \mathbf{F}(\mathbf{U}) = \begin{bmatrix} m \\ \frac{m^2}{\rho} + P \\ \left(\frac{m}{\rho}\right)(e+P) \end{bmatrix} \tag{9.2}$$

and where ρ is the density, u is the velocity, $m = \rho u$ is momentum, P is the pressure and e is the internal energy per unit volume. The variables are related by $e = \rho\varepsilon + \frac{1}{2}\rho u^2$, where $\varepsilon = \frac{P}{(\gamma-1)\rho}$ is the internal energy per internal mass with γ the ratio of specific heats (which is taken to be 1.4).

We study the Riemann problem which was described in chapter 8. We review here its most important characteristics. There is a shock tube with 2 gases separated by a membrane. Initially both gases are at rest and are at different pressures and densities defined by $P_4 > P_1$ and $\rho_4 > \rho_1$ where the subscript refers to the region in which the variables are defined (initially the region (4) is at the left of the membrane and region (1) is at the right of the membrane; afterwards the region (4) is the first region from the left boundary and the region (1) is the first region from the right boundary (see Fig. 9.2).

The exact solution can be found explicitly as a function of x and t (Liepmann and Roshko [132]) and its plot (numerical solution versus analytical solution) is shown in Fig. 9.2. The analytical expression of the exact solution is presented in chapter 10. The solution has several distinct regions: a region of low pressure and density; an area between shock and contact discontinuity; an area between contact discontinuity and rarefaction wave; a rarefaction wave and a region of high pressure and density.

9.2 Tangent linear system approach for the sensitivity computation

We consider a symbolic form for a time-dependent system of equations

$$\frac{\partial \mathbf{X}}{\partial t} = N(\mathbf{X}) \tag{9.3}$$

Then the perturbed solution $(\mathbf{X}(t) + \delta\mathbf{X}(t))$ satisfies the equation

$$\begin{aligned} \frac{\partial(\mathbf{X}(t) + \delta\mathbf{X}(t))}{\partial t} &= N(\mathbf{X}(t) + \delta\mathbf{X}(t)) = \\ N(\mathbf{X}(t)) &+ \frac{\partial N}{\partial \mathbf{X}}(\mathbf{X}(t))\delta\mathbf{X}(t) + O(\delta\mathbf{X}(t)) \end{aligned}$$

where $\frac{\partial N}{\partial \mathbf{X}}$ is the Jacobian of the nonlinear function N with respect to the variables $\mathbf{X}$. Upon retaining only the first order terms in $\delta\mathbf{X}$ the previous equation becomes

$$\frac{\partial \delta\mathbf{X}(t)}{\partial t} = \frac{\partial N}{\partial \mathbf{X}}(\mathbf{X}(t))\delta\mathbf{X}(t) \tag{9.4}$$

To determine the sensitivity with respect to a parameter α we differentiate equation (9.3) and assuming that we can interchange the order of differentiation we obtain

$$\frac{\partial}{\partial t}\left(\frac{\partial \mathbf{X}}{\partial \alpha}\right) = \frac{\partial N(\mathbf{X})}{\partial \mathbf{X}}\frac{\partial \mathbf{X}}{\partial \alpha} \tag{9.5}$$

which implies that the sensitivity $\frac{\partial \mathbf{X}}{\partial \alpha}$ with respect to the parameter α satisfies also the tangent linear equation (9.4).

This provides the rationale for the numerical computation of the sensitivity using the tangent linear model.

9.3 Linearization of the Euler equations

The following derivation follows Godlewski and Raviart [71].

Given a solution of (9.1), called *basic solution*, we study the behavior in time of solutions of the linear hyperbolic system obtained by linearizing (9.1) at the basic solution. Since the basic solution is discontinuous, the linearized system has discontinuous coefficients and it is not well posed in any class of functions. The solution of the linearized system consists of the sum of a function and a measure caused by the discontinuity of the basic solution.

Let $\mathbf{U} = \mathbf{U}(x,t)$ be the basic solution and the first order perturbation $\mathbf{V}$. We construct $\mathbf{U}^\epsilon$ which satisfies:

$$\begin{aligned} \mathbf{U}^\epsilon &= \mathbf{U}(x,t) + \epsilon\mathbf{V}(x,t) \\ \mathbf{U}^\epsilon(x,0) &= \mathbf{U}(x,0) + \epsilon\mathbf{V}(x,0) = \mathbf{U}_0(x) + \epsilon\mathbf{V}_0(x) \end{aligned}$$

with $\epsilon > 0$ a small parameter.

The first order perturbation $\mathbf{V} = \mathbf{V}(x,t)$ is solution of the linearized problem

$$\frac{\partial \mathbf{V}}{\partial t} + \frac{\partial}{\partial x}(\mathbf{J}(\mathbf{U})\mathbf{V}) = 0 \tag{9.6}$$
$$\mathbf{V}(x,0) = \mathbf{V}_0(x)$$

where $\mathbf{J}(\mathbf{U})$ denote the Jacobian of $\mathbf{F}(\mathbf{U})$.

The basic solution $\mathbf{U}$ presents a discontinuity along the line $\mathbf{L} = \{(x,t), x = \Phi(t), t \geq 0\}$ where the function $\Phi(t)$ is determined by the location of the shock.

In chapter 10 we derive the shock location as given by:

$$\Phi(t) = \left(\frac{\gamma P_1}{\rho_1}\right)^{\frac{1}{2}} \left(\frac{\gamma - 1}{2\gamma} + \frac{\gamma + 1}{2\gamma}\frac{P_2}{P_1}\right)^{\frac{1}{2}} + x_0$$

where the subscripts refer to the corresponding region in which the variables P_1, P_2 and ρ_1 are defined and x_0 is the initial position of the diaphragm (at $t = 0$)).

$\mathbf{U}$ presents at most weak discontinuities outside the line $\mathbf{L}$. Although for t small enough the linearized problem retains the same characteristics as the nonlinearized model, the perturbed solution $\mathbf{U}^\epsilon$ presents a discontinuity along a different line

$$\mathbf{L}^\epsilon = \{(x,t), x = \Phi^\epsilon(t) = \Phi(t) + \epsilon\Psi(t), t \geq 0\}$$

and at most weak discontinuities outside $\mathbf{L}^\epsilon$.

We introduce the equation of the front of the discontinuities as one of the unknowns and we use a change of variables to reduce the problem to a fixed domain

$$\hat{x} = x - \Phi^\epsilon(t) \tag{9.7}$$

$$\hat{\mathbf{U}}^\epsilon(\hat{x},t) = \mathbf{U}^\epsilon(\hat{x} + \Phi^\epsilon(t), t) \tag{9.8}$$

The function $\mathbf{U}$ is now discontinuous along the fixed line $\hat{x} = 0$ and is the solution of the Cauchy problem

$$\frac{\partial \hat{\mathbf{U}}^\epsilon}{\partial t} + \frac{\partial}{\partial \hat{x}}(F(\hat{\mathbf{U}}^\epsilon) - \frac{\partial \Phi^\epsilon}{\partial t}\hat{\mathbf{U}}) = 0 \tag{9.9}$$
$$\mathbf{U}^\epsilon(\hat{x},0) = \mathbf{U}_0(\hat{x} + \Phi^\epsilon(0)) + \epsilon\mathbf{V}_0(\hat{x} + \Phi^\epsilon(0))$$

Moreover $\hat{\mathbf{U}}^\epsilon$ satisfies the Rankine-Hugoniot jump relations across $\hat{x} = 0$

$$[F(\hat{\mathbf{U}}^\epsilon)] = \frac{\partial \Phi^\epsilon}{\partial t}[\hat{\mathbf{U}}^\epsilon] \tag{9.10}$$

Recalling that

$$\hat{\mathbf{U}}(\hat{x},t) = \mathbf{U}(\hat{x} + \Phi(t), t)$$
$$\hat{\mathbf{U}}^\epsilon = \hat{\mathbf{U}} + \epsilon\hat{\mathbf{U}} + \cdots$$
$$\Phi^\epsilon = \Phi + \epsilon\Psi$$

we obtain that the pair $(\hat{\mathbf{V}}, \Psi)$ satisfies the linearized equations and the Rankine-Hugoniot relation:

$$\frac{\partial \hat{\mathbf{V}}}{\partial t} + \frac{\partial}{\partial \hat{x}}\left(\left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\hat{\mathbf{V}} - \frac{\partial \Psi}{\partial t}\hat{\mathbf{U}}\right) = 0 \tag{9.11}$$
$$\hat{\mathbf{V}}(\hat{x},0) = \mathbf{V}_0(\hat{x}) + \Psi(0)\frac{d\mathbf{U}_0}{dx}(\hat{x})$$
$$\left[\left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\hat{\mathbf{V}}\right] = \frac{\partial \Psi}{\partial t}[\hat{\mathbf{U}}]$$

Let us define $\bar{\mathbf{V}}(\hat{x},t) = \hat{\mathbf{V}}(\hat{x},t) - \Psi(t)\dfrac{\partial \hat{\mathbf{U}}}{\partial \hat{x}}(\hat{x},t)$.

The following relation

$$\frac{\partial \bar{\mathbf{V}}}{\partial t} + \frac{\partial}{\partial \hat{x}}\left(\left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\bar{\mathbf{V}}\right) = 0$$

is valid, since

$$\begin{aligned}
& \frac{\partial \bar{\mathbf{V}}}{\partial t} + \frac{\partial}{\partial \hat{x}}\left(\left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\bar{\mathbf{V}}\right) \\
= \; & \frac{\partial \hat{\mathbf{V}}}{\partial t} - \frac{\partial \Psi}{\partial t}\frac{\partial \hat{\mathbf{U}}}{\partial \hat{x}} - \Psi\frac{\partial^2 \hat{\mathbf{U}}}{\partial t \partial \hat{x}} \\
+ \; & \frac{\partial}{\partial \hat{x}}\left(\left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\hat{\mathbf{V}}\right) - \frac{\partial}{\partial \hat{x}}\left(\left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\Psi\frac{\partial \hat{\mathbf{U}}}{\partial \hat{x}}\right) \\
= \; & -\frac{\partial}{\partial \hat{x}}\left(\Psi\left[\frac{\partial \hat{\mathbf{U}}}{\partial t} + \left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\frac{\partial \hat{\mathbf{U}}}{\partial \hat{x}}\right]\right) = 0
\end{aligned}$$

In conclusion, he pair $(\bar{\mathbf{V}}, \Psi)$ satisfies the following equations and jump condition:

$$\begin{aligned}
\frac{\partial \bar{\mathbf{V}}}{\partial t} + \frac{\partial}{\partial \hat{x}}\left(\left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\bar{\mathbf{V}}\right) &= 0 \qquad (9.12) \\
\bar{\mathbf{V}}(\hat{x},0) &= \mathbf{V}_0(\hat{x}) \\
\left[\left(\mathbf{J}(\hat{\mathbf{U}}) - \frac{\partial \Phi}{\partial t}\right)\bar{\mathbf{V}}\right] &= \frac{\partial \Psi}{\partial t}[\hat{\mathbf{U}}] + \Psi\frac{\partial \hat{\mathbf{U}}}{\partial t}
\end{aligned}$$

The equations (9.12) have a unique solution (Godlewski [71]).

Finally, we can define the solution $\mathbf{V}$ of the system (9.6) as the sum of a function and a measure whose support is $\mathbf{L}$

$$\mathbf{V}(x,t) = \bar{\mathbf{V}}(x - \Phi(t), t) - \Psi[\mathbf{U}]\delta_{\mathbf{L}} \qquad (9.13)$$

where $\delta_{\mathbf{L}}$ is the Dirac measure with support $\mathbf{L}$.

9.4 L^2 estimates for the solution of the linearized Euler equations

L^2 estimates for the solution of the linearized system are obtained following the approach of Metivier [143]. Let $\Omega = \mathbf{R} \times [0,\infty]$ and $\omega = \{x = 0\}$ the boundary of Ω. We define $L^2_\eta = e^{\eta t}L^2$ and $H^1_\eta = e^{\eta t}H^1$.

In these spaces we consider the norms

$$||u||^2_{L^2_\eta} = \int_\Omega e^{-2\eta t}|u(x)|^2 dx = ||e^{-\eta t}u||^2_{L^2}$$

$$||u||^2_{H^1_\eta} = ||e^{-\eta t}u||^2_{H^1} \tag{9.14}$$

The solution $\left(\bar{\mathbf{V}}, \Psi\right) \in H^1_\eta(\Omega) \times H^1_\eta(\omega)$ of (9.12) satisfies the estimate

$$\eta||\bar{\mathbf{V}}||^2_{L^2_\eta(\Omega)} + ||\bar{\mathbf{V}}|_{x=0}||^2_{L^2_\eta(\omega)} + ||\nabla\Psi||^2_{L^2_\eta(\omega)} \leq \frac{C}{\eta}||e^{-\eta t}\mathbf{F}||^2_0 \tag{9.15}$$

where C is a constant, $\eta \geq \eta_0$ with η_0 given and $||u||_{L^2_\eta} = ||e^{-\eta t}u||_0$, with $||u||_0$ being the usual norm in L^2.

9.5 Numerical model using Adaptive Mesh Refinement

To solve the Riemann problem we chose a code written by S. Li [130] which employs a method of adaptive mesh refinement in conjunction with a Riemann solver of Roe-type (Leveque [129]). The numerical solution is in very good agreement with the analytical solution and this eliminates a major source of errors in the numerical computation of the sensitivities.

We describe the method of adaptive mesh refinement by taking the grid in Fig. 9.1 as an example.

The grid has three refinement levels (L_0, L_1, L_2) (the order is from the coarsest level to the finest level) at time t_m. Let us suppose that the corresponding time step sizes, for the corresponding refinement levels, are as follows:

$$\Delta t_0 = t_{m+1} - t_m, \quad \Delta t_1 = \frac{1}{2}\Delta t_0, \quad \Delta t_2 = \frac{1}{4}\Delta t_0$$

The AMR method consists of the following steps:

- Step 1. We start from the coarsest level which advances one time step Δt_0.
- Step 2. L_1 advances its corresponding time step Δt_1 and the boundary conditions are obtained from L_0.
- Step 3. L_2 advances two time steps Δt_2. We update the solution on L_1 with the solution on L_2 which is more accurate.
- Step 4. L_1 advances one more corresponding time step Δt_1
- Step 5. L_2 advances two more time steps Δt_2. Now we update the solutions on L_0 and L_1 respectively with the more accurate solution on L_2.
- Step 6. We readapt the mesh (based on an "a posteriori" error estimate) and generate a new hierarchical grid.

The hierarchical grid data structure $G = |n|G_1|G_2| \cdots |G_n|$ contains the number n of levels of the grid and pointers to the grid on each of the lower levels. The data structure for the grid on the $i-$th level $G_i = |m_i|p_i|G_{i,1}|p_2|G_{i,2}| \cdots |p_{m_i}|G_{i,m_i}|$ contains the number of patches m_i, information about the $j-$th patch on the $i-$th level (denoted $G_{i,j}$) and a pointer p_j to the parent patch for $G_{i,j}$ to facilitate operations between the coarse and fine grids.

The patch has the following attributes: level of the current patch P, integration time and time step size for P, number of ghost boundary points, number of grid points, grid index andphysical grid location for each point, spatial step length, solution values, pointers to parent patches (only one in 1-D), refinement ratio between the parent patch and P and pointers to siblings (in 2-D and 3-D).

The solution is obtained using an AMR recursive integration algorithm:

```
INTEGRATE($level$); BEGIN
• Let $maxlevel$ the maximum level allowable and $flevel$ the finest level existing
• Remesh Stage($level$):
    – $flevel = \max(flevel + 1, maxlevel)$
    – WHILE ($flevel - 1$ needs no refining) decrease $flevel$ by 1
    – FOR $slevel = flevel - 1$ DOWNTO $level$ DO Refine($slevel$)
        ∗ Select($slevel$)
        ∗ Expand($slevel$)
        ∗ Cluster($slevel$)
    – FOR $slevel = level$ UPTO $flevel - 1$ DO Regrid($slevel + 1$)
    – WHILE ($flevel < maxlevel$ AND $flevel$ needs refining) DO
        ∗ Refine($flevel$)
        ∗ Regrid($flevel + 1$); Increase $flevel$ by 1
• Advance Solution($level$)
    – Boundary Collection($level$)
    – Advance($level$)
• Recursive Stage($level$)
    – IF $level \neq flevel$ THEN FOR $r = 0$ TO $\Delta t(level)/\Delta t(level + 1)$ DO
        ∗ INTEGRATE($level + 1$)
• Project Solution($level, level + 1$)
END
```

The **Remesh** stage is split into two main processes: first readapt the available grid and then refine to generate the new finer grid. The refinement is divided into **Selection** (to flag the inaccurate points which needs refining), **Expansion** (to add buffer zones around the flagged region) and **Clustering** (to group the flagged points into clusters).

There may be features which appear in the finer levels which would not be captured if we start this first process of readaptation from the coarsest grid. For this reason we initialize the mesh refinement on the finest level available. The **Regridding** step (in which we define the solution values for the readapted grid) starts from the coarsest level available and we continue to **Refine** and **Regrid** if the finest level available does not reach the maximum level allowable.

In the **Select** step we flag the grid points that need to be refined. The monitor function to be used has the formula:

$$Mon(i) = \max_j \frac{W(j)}{Umax(j) \cdot TOL}(|\Delta x^2 u^j_{xx}(i)|) \tag{9.16}$$

where $W(j) \in (0, 1)$ indicates the relative importance of a PDE component, $Umax(j)$ denotes the approximate maximum absolute value for each component u^j of the solution at grid point x_i and TOL is a spatial error tolerance.

A level refinement is initialized if there is a point where $Mon(i) > 1.0$ and then all the grid points with $Mon(i) > 0.5$ are flagged (if the current level grid has a grandparent, those points are also flagged).

The **Expand** step is performed not to let escape the most interesting features of the solution escape the refinement region. We add buffer zones to the flagged points every k time steps.

During the **Cluster** step adjacent flagged points are grouped together and patches which are within two buffer zones are joined together.

Regridding the $(l+1)$ level includes computing the physical location for each fine grid and obtaining the solution for the new grid. Conservative interpolation is combined with the Minmod limiter. A new refinement is partially or completely contained in an existing patch. This refinement will preserve accuracy in the domains with discontinuities due to overlapping regions between the old and the new grid.

During the **Advance** stage the solver advances the solution for all the patches in a level l one time step. The boundary values are obtained from the level $l-1$ (**Boundary Collection**). The external boundaries are given. The internal boundaries (needed by the patches) are computed using linear interpolation between the values for the internal boundaries from the parent coarse grid at the forward time t_{n+1} and the boundary values at t_n on the finer grid.

If necessary the **Recursive Stage** is performed to advance the solution on the next level $l+1$ one time step.

After integrating the finest grid T time steps the finest grid reaches the same time level as the coarser grid and the coarser grid starts integration again. The solution is updated using **Projection**: the more accurate values of the solution on the finer level replace the values of the solution on the coarser level when they coincide.

9.6 Numerical considerations

One cannot differentiate the flow across the shock or the contact discontinuity since the flow is not even continuous there. As one differentiates across the phenomena Dirac *delta* functions will appear at these locations. The flow is also not differentiable (although continuous) at the edges of the rarefaction wave. Differentiation across the edges of that wave result in jump discontinuities in the sensitivities.

However the flow solution can be differentiated within each of the five regions. The numerical derivatives at the edges make sense if constructed limits (left or right) of the derivatives inside the five regions.

The tangent linear model is obtained at the level code, being the discrete equivalent of the linearization around the basic state. We computed the sensitivity of the flow variables (pressure, density and velocity) with respect to an initial parameter (the high pressure initial condition at the left of the membrane).

The numerical sensitivities show spikes at locations where the analytic derivatives do not exist. They approximate relatively well the phenomenom of Dirac measure at the edges (where the analytical solution is not differentiable). This is to be expected since the purpose of this research is to obtain numerical sensitivities which are *as close as possible* to the analytical sensitivities.

Our results (see Fig. 9.3-9.5) approximate very well the exact sensitivities in the five regions. We compared them with previously obtained numerical results (Gunzburger [86] computed the numerical sensitivity using finite differences, the sensitivity equation and automatic differentiation). Our results show improvement both inside the five regions and at the edges of these regions where the flow is not differentiable. We think that this improvement is mainly due to the implementation of the tangent linear model derived from a forward model with adaptive mesh refinement.

First we discuss our results at the locations where the flow is continuous (i.e., inside the five regions). Both our numerical sensitivities and the numerical values presented by Gunzburger in [86] practically coincide to the analytical sensitivities on these regions.

At the edges of the five regions the situation is different. We have non differentiable points there, which result in spikes in the graph of the analytical sensitivities. The numerical sensitivities attempt to approximate these spikes. The main difference between our results and the results in [86] can be seen around the location of the shock wave. The amplitude of the numerical spike in our case is 1.15 for the derivative of the velocity, 0.5 for the derivative of the pressure and 0.35 for the derivative of the density (compared to 3.5, 0.85 and respectively 0.55 in [86]).

We chose the adaptive mesh refinement coupled with a Riemann solver as the forward model to eliminate as much as possible the errors propagating from solving numerically the discontinuities. The consequence of this choice is a much better approximation of the sensitivities at the location of discontinuities.

Our experience with tangent linear models in higher dimensions (although the application did not involve non smooth functions: Homescu et al. [101]) suggests the possibility of application of the numerical methodology presented here for spatial higher dimensions. Future directions of research include the application of this methodology for problems in 2-D, where we expect a decrease in the numerical accuracy of sensitivity computation. A possible remedy in order (Dadone et al. [44]) to alleviate this problem is to apply a smoother to the sensitivities after they were computed using the tangent linear model.

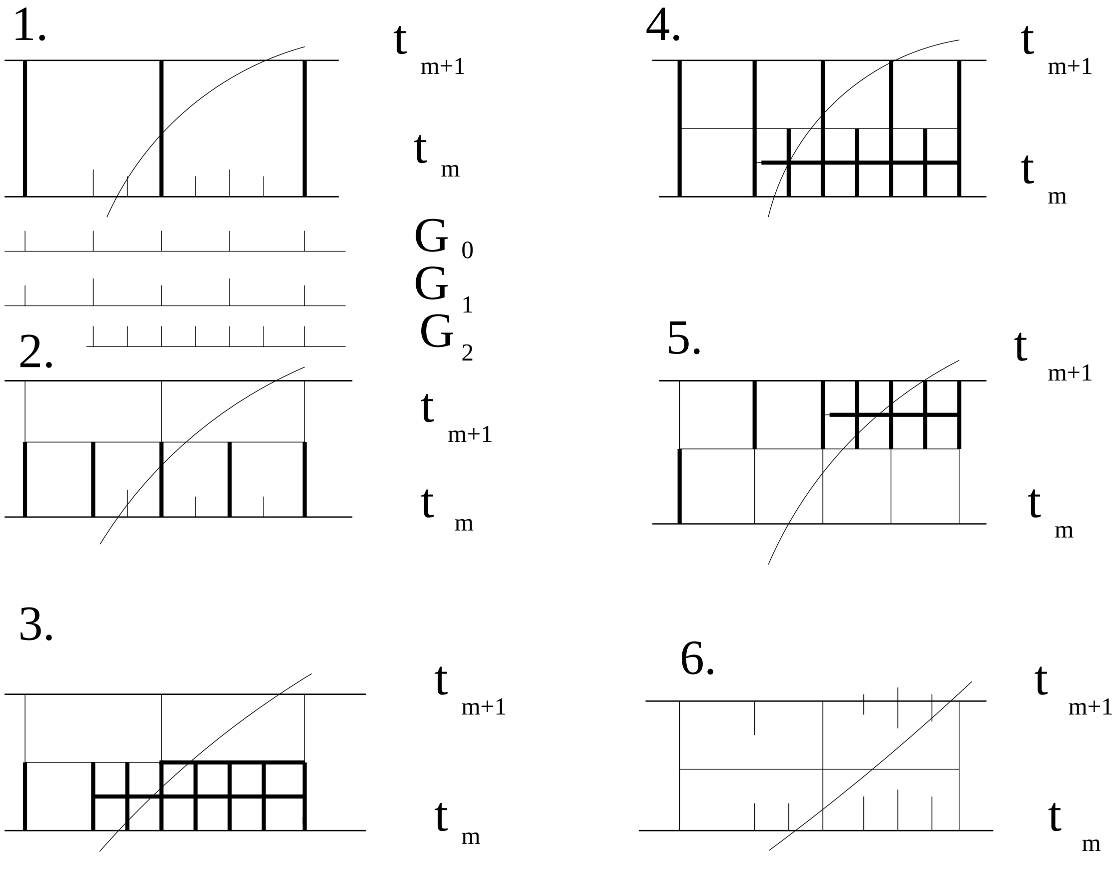
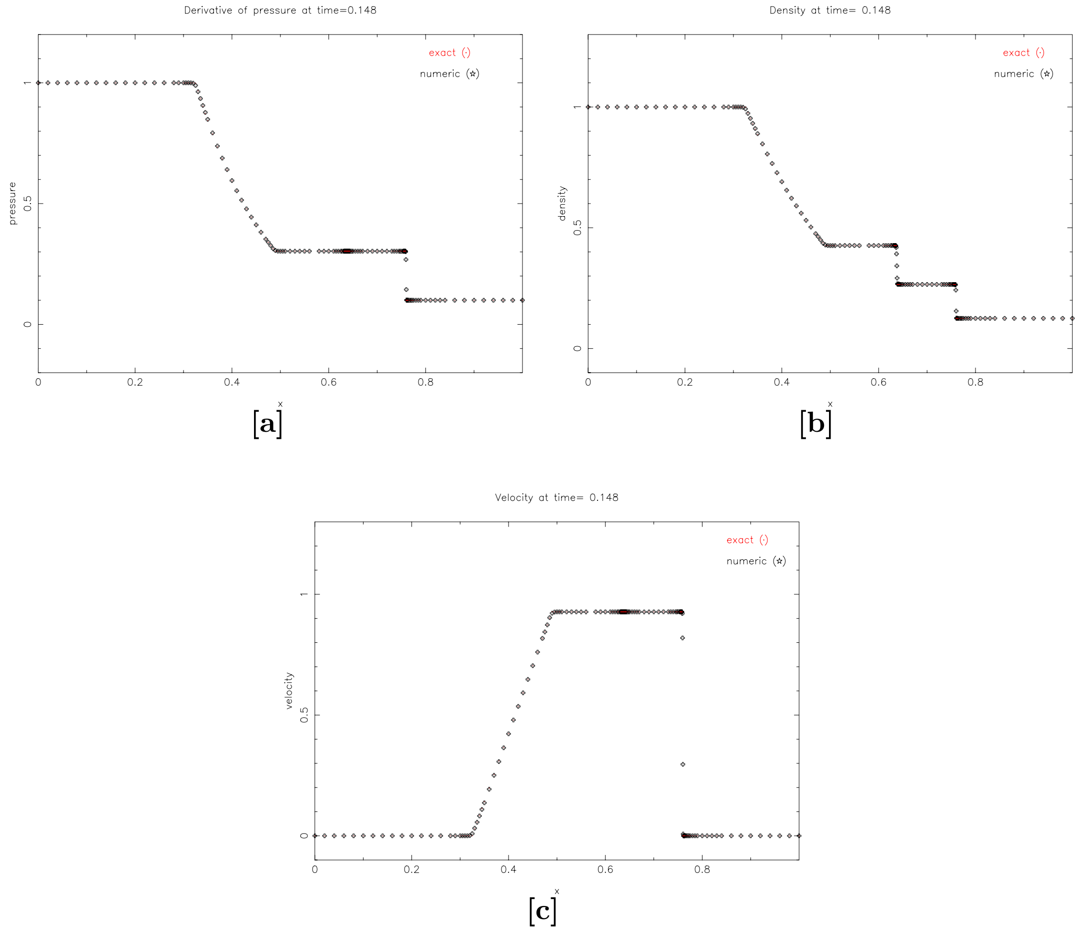


Figure 9.1. The process of adaptive mesh refinement



[a]

[b]

[c]

Figure 9.2. Exact solution of the shock-tube problem: numerical and exact values for **[a]** pressure, **[b]** density and **[c]** velocity.

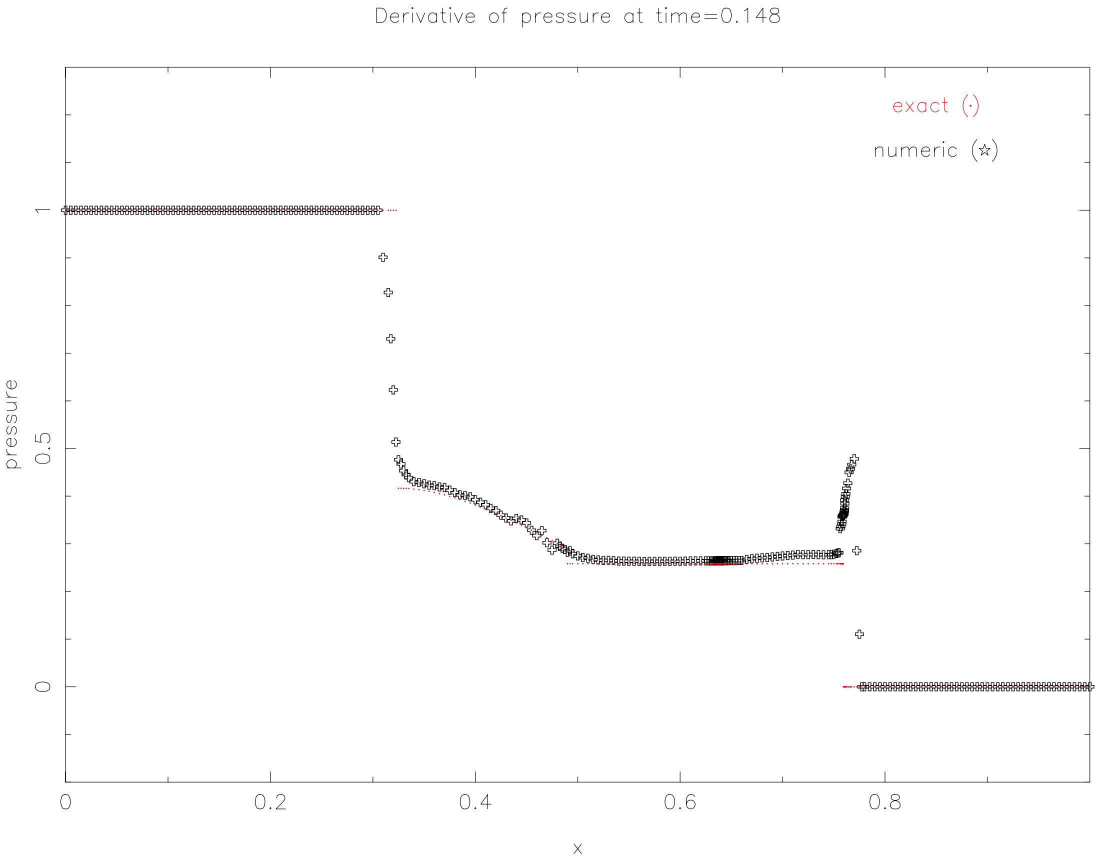


Figure 9.3. Sensitivity with respect to the high initial pressure: numerical and exact values for pressure

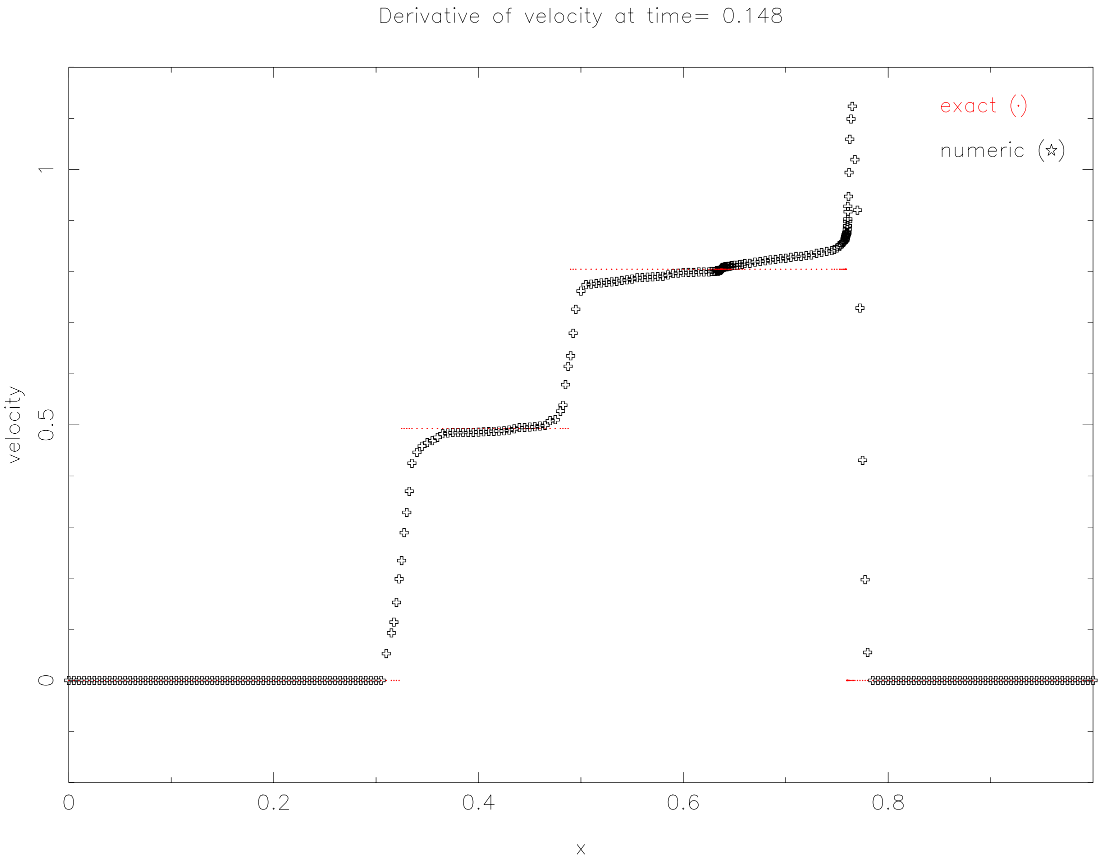


Figure 9.4. Sensitivity with respect to the high initial pressure: numerical and exact values for velocity

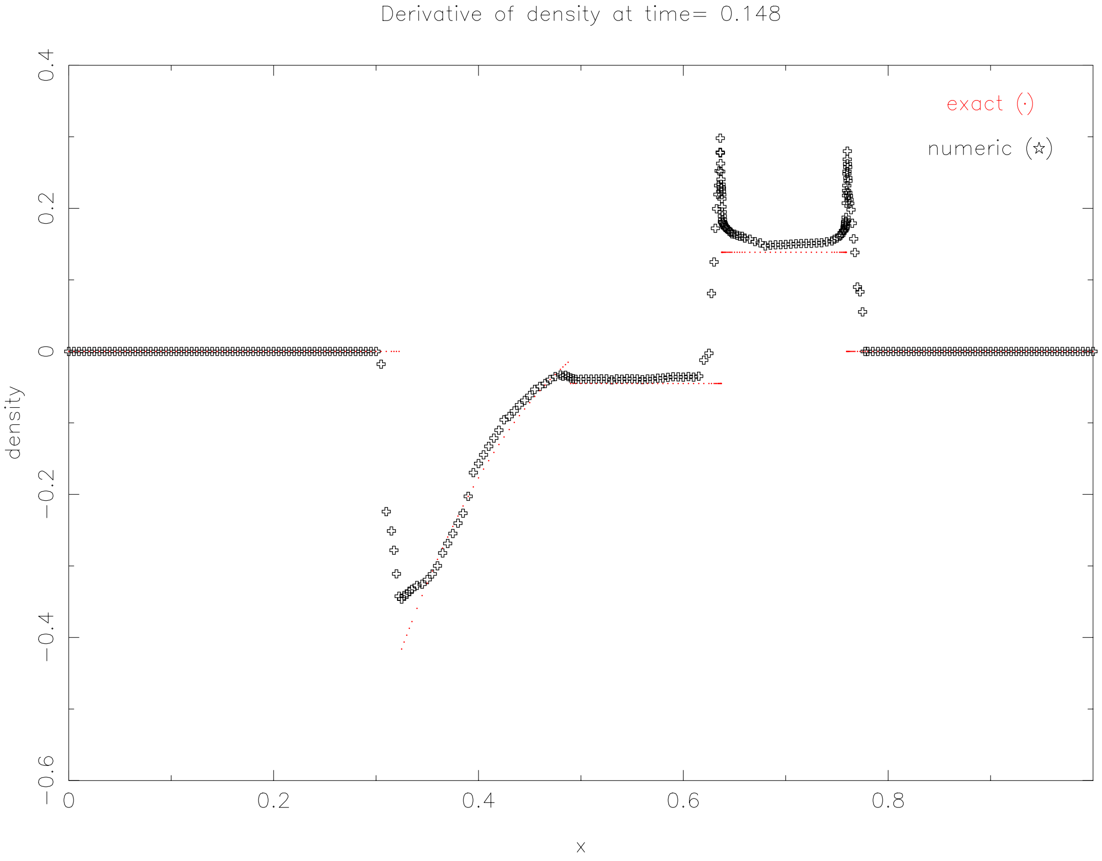


Figure 9.5. Sensitivity with respect to the high initial pressure: numerical and exact values for density

CHAPTER 10

OPTIMAL CONTROL OF FLOW WITH DISCONTINUITIES

10.1 Governing equations

Let us remind you the conservation law form of the one-dimensional unsteady equations of gas dynamics (Euler equations):

$$\mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = 0 \tag{10.1}$$

where

$$\mathbf{U} = \begin{bmatrix} \rho \\ m \\ e \end{bmatrix}, \qquad \mathbf{F}(\mathbf{U}) = \begin{bmatrix} m \\ \frac{m^2}{\rho} + P \\ \left(\frac{m}{\rho}\right)(e+P) \end{bmatrix} \tag{10.2}$$

ρ is the density, u is the velocity, $m = \rho u$ is momentum, P is the pressure and e is the internal energy per unit volume. The variables are related by $e = \rho\varepsilon + \dfrac{1}{2}\rho u^2$, where $\varepsilon = \dfrac{P}{(\gamma-1)\rho}$ is the internal energy per internal mass with γ the ratio of specific heats (which is taken to be 1.4).

We consider the Riemann problem of the Euler equations which, as mentioned in earlier chapters, corresponds to the "shock-tube problem". We review here its principal characteristics: a tube, filled with gas, is initially divided by a membrane into two sections. The gas has a higher density and pressure in one half of the tube than in the other half, with zero velocity everywhere. The initial conditions for density, velocity and pressure are similar to the values for the Sod shock-tube problem [177]:

$$\rho_{left} = 1.0 > \rho_{right} = 0.125, \;\; u_{left} = u_{right} = 0.0, \;\; p_{left} = 1.0 > p_{right} = 0.1$$

where the subscripts *left* and *right* correspond to the initial position with respect to the membrane. At time $t = 0$ the membrane is suddenly removed and the gas is allowed to flow. We expect a net motion in the direction of lower pressure. Assuming uniform flow across the tube, there is variation in only one direction and the 1-D Euler equations apply. One should calculate the flow variables: pressure, density and velocity as a function of time and space.

The solution of this Riemann problem for Euler equations consists of 5 distinct regions (see Fig. 10.13). The description of these regions follows with the corresponding region index in the parentheses: low pressure and density region (region 1), area between shock and contact discontinuity (region 2), area between contact discontinuity and rarefaction wave (region 3), rarefaction wave region (region R), high pressure and density region (region 4).

The exact solution can be found explicitly as a function of x and t (Liepmann and Roshko [132]). It is given by the following equations (the indices $1, 2, 3, 4$ and R are related to the above mentioned 5 regions):

$$\begin{bmatrix} P \\ \rho \\ u \end{bmatrix} = \begin{cases} \begin{bmatrix} P_4 \\ \rho_4 \\ u_4 \end{bmatrix} = \begin{bmatrix} P_{high} \\ \rho_{high} \\ u_{high} \end{bmatrix}, & x < -a_4 t + c \\ \begin{bmatrix} P_R \\ \rho_R \\ u_R \end{bmatrix} = \begin{bmatrix} P_R \\ \rho_R \\ u_R \end{bmatrix}, & -a_4 t + c \le x \le \left(\frac{\gamma_4+1}{2} u_3 - a_4\right) t + c \\ \begin{bmatrix} P_3 \\ \rho_3 \\ u_3 \end{bmatrix} = \begin{bmatrix} P_2 \\ \rho_2 \\ u_2 \end{bmatrix}, & \left(\frac{\gamma_4+1}{2} u_3 - a_4\right) t + c \le x \le u_2 t + c \\ \begin{bmatrix} P_2 \\ \rho_2 \\ u_2 \end{bmatrix} = \begin{bmatrix} \phi \\ \rho_2 \\ u_2 \end{bmatrix}, & u_2 t + c < x < a_1 \left(\frac{\gamma_1 - 1}{2\gamma_1} + \frac{\gamma_1 + 1}{2\gamma_1} \frac{P_2}{P_1}\right)^{\frac{1}{2}} t + c \\ \begin{bmatrix} P_1 \\ \rho_1 \\ u_1 \end{bmatrix} = \begin{bmatrix} P_{low} \\ \rho_{low} \\ u_{low} \end{bmatrix}, & a_1 \left(\frac{\gamma_1 - 1}{2\gamma_1} + \frac{\gamma_1 + 1}{2\gamma_1} \frac{P_2}{P_1}\right)^{\frac{1}{2}} t + c \le x \end{cases}$$

where $a_i^2 = \dfrac{\gamma_i P_i}{\rho_i}$, $\gamma_i = \gamma$ for $i = 1, \ldots, 4$
and where ϕ is given implicitly by

$$\frac{P_4}{P_1} = \frac{\phi}{P_1}\left(1 - \frac{(\gamma_4 - 1)(a_1/a_4)(\phi/P_1 - 1)}{\sqrt{2\gamma_1}\sqrt{2\gamma_1 + (\gamma_1 + 1)(\phi/P_1 - 1)}}\right)^{\dfrac{-2\gamma_4}{\gamma_4 - 1}} \tag{10.3}$$

The remaining variables : ρ_2, u_2 and ρ_3 are given by

$$\rho_2 = \rho_1 \frac{P_2}{P_1}\left(\frac{1 + \frac{\gamma_1 - 1}{\gamma_1 + 1}\frac{P_1}{P_2}}{1 + \frac{\gamma_1 - 1}{\gamma_1 + 1}\frac{P_2}{P_1}}\right) \tag{10.4}$$

$$u_2 = a_1\left(\frac{P_2}{P_1} - 1\right)\sqrt{\frac{2\gamma_1}{(\gamma_1 + 1)\frac{P_2}{P_1} + (\gamma_1 - 1)}} \tag{10.5}$$

$$\rho_3 = \rho_4\left(\frac{P_3}{P_4}\right)^{\frac{1}{\gamma_4}} \tag{10.6}$$

and, in the rarefaction wave, the quantities P_R, ρ_R and u_R are given by

$$P_R = P_4\left(1 - \frac{\gamma_4 - 1}{2}\frac{u_R}{a_4}\right)^{\frac{2\gamma_4}{\gamma_4 - 1}} \tag{10.7}$$

$$\rho_R = \rho_4\left(1 - \frac{\gamma_4 - 1}{2}\frac{u_R}{a_4}\right)^{\frac{2}{\gamma_4 - 1}} \tag{10.8}$$

$$u_R = \left(\frac{u_3 - u_4}{\frac{\gamma_4+1}{2}u_3}\right)\left(\frac{x - c}{t}\right) + \frac{a_4u_3 - (a_4 - \frac{\gamma+4+1}{2}u_3)u_4}{\frac{\gamma_4+1}{2}u_3} \tag{10.9}$$

The subscripts for the variables $u_2, p_R, \ldots$ match the corresponding region of the solution in which they are located (for example, p_R is the value of the pressure in the rarefaction region).

We also present the formula for the physical entropy. The entropy is necessary while selecting the solution of the shock-tube problem in the weak sense. The correct weak solution should satisfy the entropy condition, which states that the entropy of fluid particles does not decrease. We are employing the following formula for the entropy S (Wesseling [204]):

$$S = c_V \ln\left(\frac{p}{\rho^\gamma}\right) \tag{10.10}$$

where c_V is the specific heat at constant volume, p is the pressure, ρ is the density and γ the ratio of specific heats.

10.2 Description of the numerical models: AVM and HRM

The main difficulties encountered when solving numerically the shock-tube problem of gas dynamics (and, in general, for any problem which has a non smooth solution) appear in the regions of discontinuities. The numerical solution may be *s*moothed in those regions (e.g., due to introduction of a dissipation term) or the discontinuities ca be captured in a sharper way (using high-resolution methods). For this reason we chose one numerical model from each of the above mentioned categories: namely a model with artificial viscosity (**AVM**) and a high-resolution model (**HRM**) with a Riemann solver.

As a footnote we mention that for very accurate numerical solutions adaptive mesh refinement **AMR** may be used in conjunction with Riemann solvers (e.g., Leveque [128] for Euler equations). Our experience with a model **AMR** in the framework of sensitivity analysis for discontinuous flows was presented in chapter 9 (also in Homescu and Navon [100]).

Our research aims to perform optimal control of flow with discontinuities using either smooth or non smooth optimization techniques for minimizing the cost functional. The minimization requires availability of either the gradient or of subgradients for the cost functional (with respect to the control variables) obtained using the adjoint model derived from the forward model (either **AVM** or **HRM** models).

10.2.1 The numerical model with artificial viscosity AVM

The **AVM** model (by T.J. Cowan [41]) uses finite elements which are piecewise constant in time and piecewise linear in space. The elements are discontinuous in time but continuous in space. By using discontinuous discretization in time we were able to march sequentially through time and solve for only a fraction of the total solution at one time. To improve the stability of the method a least-squares operator is added to the basic Galerkin formulation. In order to obtain non-oscillatory approximations to discontinuities, discontinuity-capturing operators have been developed within the framework of this modified discontinuous Galerkin/least squares method (Shakib et al. [174]).

An artificial viscosity term (included to stabilize the numerical solution) has the effect of spreading flow discontinuities over several computational cells. The method employs a high-order scheme for the smooth regions of the flow combined with a low-order solution which is employed near the discontinuities.

The combination, described in Lohner et al. [163], is based on the generalization of flux-corrected transport (**FCT**) algorithm developed by Zalesak [208]. **FCT** combines a high-order scheme with a low-order scheme. The high-order scheme is employed in regions where the variables under consideration vary smoothly (so that a Taylor expansion makes sense), whereas in those regions where the variables vary abruptly the schemes are combined, in a conservative manner, in an attempt to ensure a monotonic solution.

Let us assume that the temporal discretization of the Euler equations yields

$$U^{n+1} = U^n + \Delta U \tag{10.11}$$

where ΔU is the increment of unknowns obtained for a given scheme at time $t = t^n$. Our aim is to obtain a ΔU of as high an order as possible without introducing overshoots. To this end we rewrite the equation (10.11) as

$$U^{n+1} = U^n + \Delta U^{low} + (\Delta U^{high} - \Delta U^{low}) = U^{low} + (\Delta U^{high} - \Delta U^{low}) \tag{10.12}$$

where ΔU^{high} and ΔU^{low} denote the increments obtained by some high- and low- order scheme and U^{low} is the monotone, ripple-free solution at time $t = t^{n+1}$ of the low-order scheme.

The idea behind **FCT** is to limit the second term on the right-hand side of equation (10.12) in such a way that no new over/undershoots are created. A further constraint, given by the conservation law itself, must be also taken into account: strict conservation on the discrete level should be maintained. The simplest way to guarantee this for node-centered schemes is by constructing schemes for which the sum of the contributions of each individual element (cell) to its surrounding nodes vanishes (“ all that comes in goes out”).

FCT consists of the following steps:

1. Compute the *low-order contribution* (LEC) from some low-order scheme guarantee to give monotonic results for the problem at hand;

2. Compute the *high-order contribution* (HEC) given by some high-order scheme;

3. Define the *antidiffusive element contributions* (AEC):

   $$AEC = HEC - LEC$$

4. Compute the updated low-order solution:

   $$U^{low} = U^n + \sum_{elem} LEC = U^n + \Delta U^{low}$$

5. Limit ("correct") the AEC so that U^{n+1} (as computed in step 6 below) is free of extrema:
$$AEC^{corr} = \lambda * AEC \qquad 0 \leq \lambda \leq 1$$

6. Apply the limited AEC:
$$U^{n+1} = U^{low} + \sum_{elem} AEC^{corr}$$

The high-order scheme chosen was the consistent-mass Taylor-Galerkin while the low-order scheme employed was the lumped-mass Taylor-Galerkin scheme plus diffusion.

10.2.2 Numerical high-resolution model HRM

The **HRM** model is part of the package CLAWPACK written by R. Leveque [129], [128]) which employs Roe's approximate Riemann solver (Roe [165]) combined with an entropy fix.

We present the basic ideas of the Roe solver. Let us consider a standard form of a homogeneous conservation law:

$$q_t(x,t) + f(q(x,t))_x = 0 \tag{10.13}$$

The basic algorithm depends on a Riemann solver that, for each set of data (q^L, q^R) returns a set of M_w waves W^p and speeds λ^p satisfying

$$\sum_{p=1}^{M_w} \mathcal{W}^p = q^R - q^L \equiv \Delta q$$

It also returns the left-going and right-going flux differences $\mathcal{A}^- \Delta q$ and $\mathcal{A}^+ \Delta q$ that satisfy the relationship:

$$\mathcal{A}^- \Delta q + \mathcal{A}^+ \Delta q = f(q^R) - f(q^L) \tag{10.14}$$

The Roe solver employed here consists of solving a particular linear system

$$q_t + A_i q_x = 0 \tag{10.15}$$

where A_i is the Roe matrix depending on data (q_{i-1}, q_i).

The solution consists of waves of the form $\mathcal{W}_i^p = \lambda_i^p r_i^p$ where r_i^p is the $p-$th eigenvector of A_i, which propagate with speeds λ_i^p, the corresponding eigenvalue of A_i.

The flux differences are defined as:

$$\mathcal{A}^+ \Delta q_i = \sum_{\lambda_i^p > 0} \lambda_i^p \mathcal{W}_i^p$$
$$\mathcal{A}^- \Delta q_i = \sum_{\lambda_i^p < 0} \lambda_i^p \mathcal{W}_i^p$$

Linearized Riemann problem solutions consist of discontinuous jumps only. This can be a good approximation for contacts and shocks, in that the discontinuous character of the wave is correct, although the size of the jump may not be correctly approximated by the linearized solution. Rarefaction waves, on the other hand, carry a continuous change in flow variables and, as time increases, they tend to spread. In that case the linearized approximation via discontinuous jumps is inexact. In a practical computational setup, however, linearized approximations encounter difficulties only if the rarefaction wave is *transonic*. In this case unphysical, entropy violating discontinuous waves may appear.

Roe's solver can be modified to avoid entropy violating solutions. This is usually referred to as an *entropy fix*. We employed an entropy fix for the Roe's method developed by Harten and Hyman [96], entropy fix which has widespread use. Other ways of correcting the scheme have been discussed by Roe [166] and Dubois and Mehlman [49], among others.

10.3 Existence of the solution of the optimal control problem

We solve the following optimal control problem

$$\textit{Minimize the cost functional } \boldsymbol{\mathcal{J}}(\mathbf{U}, z) \textit{ subject to } z \in \mathcal{U}_{ad} \qquad \text{(OPT)}$$

where z is the control, $\mathcal{U}_{ad}$ is the space of admissible controls and $\mathbf{U} = \mathbf{U}(z)$ is the entropy solution of the system of conservation laws (Euler 1-D equations described in the previous section):

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial F(\mathbf{U})}{\partial x} = 0 \tag{10.16}$$

$$\mathbf{U}(x, 0) = z(x) \tag{10.17}$$

with $0 \leq x \leq 1$ and $0 \leq t \leq T_W$ (T_W being the length of the assimilation window).

Since the solution of the system (10.16) may develop discontinuities after a finite time, weak solutions should be considered. An additional entropy condition must be imposed to select the "physically" relevant weak solution.

We define an entropy function $\mathcal{EF}$, for which an additional conservation law that holds for smooth solutions becomes an inequality for discontinuous solutions (Leveque [129], Godlewski and Raviart [51]). It is known that there exists a physical quantity called the entrop for the Euler equations of gas dynamics (which are employed for this research). The physical entropy is constant along particle paths in smooth flow and it jumps to higher values as the gas crosses a shock. The correct weak solution is picked out using a property of the entropy, namely that it can never jump to a lower value (a numerical version of this approach was employed for the high resolution model described in the previous section).

For the system of gas dynamics equations, which is a strictly hyperbolic symmetrizable nonlinear system of conservation laws, entropy functions can be found (e.g., Godlewski and Raviart [51], Leveque [129]).

We introduce the definition of the entropy solution, according to Godlewski and Raviart [51]:

A weak solution $\mathbf{U}$ *of (10.16)-(10.17) is called an entropy solution if* $\mathbf{U}$ *satisfies, for all entropy functions* $\mathcal{EF}$ *of (10.16) and for all test functions* $\phi \in C_0^1([0, 1] \times [0, \infty))$, $\phi \geq 0$,

$$\int_{t=0}^{\infty} \int_{x=0}^{x=1} \left(\mathcal{EF}(\mathbf{U}) \frac{\partial \phi}{\partial t} + F(\mathbf{U}) \frac{\partial \phi}{\partial x} \right) dxdt + \int_{x=0}^{x=1} \mathcal{EF}(z(x)) \phi(x, 0) dx \geq 0 \tag{10.18}$$

To derive existence results for optimal controls we follow the approach of Ulbrich [195]. His work, related to scalar laws of conservation with source terms, was extended to our case (the 1-D system of Euler equations without source terms).

For our problem the control vector $z(x)$ is:

$$z(x) = \left[\begin{array}{c} \rho(x,0) \\ m(x,0) \\ e(x,0) \end{array} \right] = \left[\begin{array}{c} \rho(x,0) \\ \rho(x,0)u(x,0) \\ \dfrac{P(x,0)}{(\gamma-1)} + \dfrac{1}{2}\rho(x,0)(u(x,0))^2 \end{array} \right]$$

with ρ the density, u the velocity, $m = \rho u$, P the pressure, γ the ratio of specific heats and e the internal energy per unit volume.

Since the control vector z is bounded (being obtained using initial values of the pressure, velocity and density) we may consider that the controls are in $\Big(L^\infty[0,1]\Big)^3$. If the control problem (OPT) is particularized to the optimal control problem for 1-D Euler equations for gas dynamics then the existence of the optimal controls is obtained as a consequence of four properties described below.

(P1) The function F, which appears in the system of conservation laws (10.16), is locally Lipschitz.

(P2) The admissible set $\mathcal{U}_{ad}$ is bounded in $\Big(L^\infty[0,1])\Big)^3$

and closed in $\Big(L^1_{loc}([0,1]\Big)^3$.

(P3) Let us we denote by $BV[(0,1)]$ the space of functions of bounded variations on the interval $(0,1)$. Based on the choice of controls the admissible set $\mathcal{U}_{ad}$ is bounded in $\Big(BV[(0,1)]\Big)^3$. The embedding $\Big(BV(\Omega)\Big)^3 -> \Big(L^1(\Omega)\Big)^3$ is compact for any open bounded set with Lipschitz-boundary $\Omega \subset (0,1)$ (Giusti [70]). Thus we obtain that $\mathcal{U}_{ad}$ is compact in $\Big(L^1_{loc}[0,1]\Big)^3$.

In our research the cost functional $\mathcal{J}$ for the optimal control problem (OPT) assumes two possible forms.

The first expression of the cost functional is

$$\mathcal{J}(\mathbf{U}, z) = \int_0^1 (\mathbf{U}(x, T_W) - \mathbf{U}^{obs}(x, T_W))^2 dx \tag{10.19}$$

with $\mathbf{U}^{obs} \in L^\infty([0,1])$ the observations distributed at assimilation time T_W.

The second cost functional is defined as

$$\mathcal{J}(\mathbf{U}, z) = \int_{t=0}^{t=T_W} \int_{x=0}^{x=1} (\mathbf{U}(x,t) - \mathbf{U}^{obs}(x,t))^2 dx dt \tag{10.20}$$

with $\mathbf{U}^{obs} \in \Big(L^\infty([0,1] \times (0,T_W))\Big)^3$ the observations at assimilation times $0 \le t \le T_W$.

We have the following property for both forms: (10.19) and (10.20)
(P4) The cost functional $\boldsymbol{\mathcal{J}}$ is (at least) sequentially lower semicontinuous.

Using the properties (P1)-(P4) one can prove that the optimal control problem (OPT) has a solution $\hat{z} \in \mathcal{U}_{ad}$ in a similar way to the proof of existence of optimal controls obtained by Ulbrich [195].

First we prove that if $\boldsymbol{\mathcal{J}}$ satisfies (P4) then

$$z \in (\mathcal{U}_{ad} \subset \left(L^1_{loc}[0,1]\right)^3) \hookrightarrow \boldsymbol{\mathcal{J}}(\mathbf{U}, z) \tag{10.21}$$

is sequentially lower semicontinuous.

Indeed let the sequence $(z^k) \subset \mathcal{U}_{ad}$ converge in $\left(L^1_{loc}[0,1]\right)^3$ to z_0. We have that $z_0 \in \mathcal{U}_{ad}$ using property (P2). We have also that $\mathbf{U}(z^k) \to \mathbf{U}(z_0)$ (Godlewski and Raviart [51]). It follows from property (P4) that

$$\underline{\lim}_{k \to \infty} \boldsymbol{\mathcal{J}}(\mathbf{U}(z^k), z^k) \geq \boldsymbol{\mathcal{J}}(\mathbf{U}(z_0), z_0)$$

which establishes the lower semicontinuity of the operator defined in (10.21).

Finally let (z^j) be a minimizing sequence for the optimal control problem (OPT). Using compactness of $\mathcal{U}_{ad}$ there exists a subsequence which converges to $\hat{\mathbf{z}} \in \mathcal{U}_{ad}$. We have proved that the operator (10.21) is sequentially lower semicontinuous, which implies that $\hat{\mathbf{z}}$ is a solution for the optimal control problem (OPT).

This concludes the proof of existence of solutions for (OPT).

10.4 Detection of discontinuities in data

In the setting of smooth minimization one may consider that, by eliminating some discontinuities from the computation of the cost functional and its gradient (or subgradient) one may obtain a function which is smoother (i.e., more appropriate for smooth optimization). Several approaches can be found in literature for the detection of discontinuities.

The discontinuity locking system (**DLS**) is employed for differential-algebraic equations (**DAE**) by Birta and Oren [19], Park and Barton [158], Mao and Petzold [139], to cite but a few. The idea for this approach (**DLS**) is to lock the function evaluator for the initial-value problem solver so that the equations evaluated are fixed while an integration step is being taken, thus presenting a smooth vector field to the solver.

The approach we present here is a modified application of a discrete regularization method proposed by Lee and Pavlidis [124].

Let $(x_i, y_i)_{i=0,\ldots,n}$ be the set of data points with $x_i < x_{i+1}$. We want to find the $n+1$ quantities z_i that minimize a combination of the discrete curvature and the discrete difference between observation and desired data:

$$\sum_{i=0}^{n} \alpha_i \left(\frac{z_{i+1} - z_i}{x_{i+1} - x_i} - \frac{z_i - z_{i-1}}{x_i - x_{i-1}} \right)^2 + \beta \sum_{i=0}^{n} (z_i - y_i)^2 \tag{10.22}$$

with $\alpha_0 = \alpha_n = 0$ and $\alpha_i = 1$ for $i = 1, \cdots, n-1$.

Differentiating (10.22) with respect to z_k and setting the derivatives to zero yields a system of $n+1$ equations with $n+1$ equations, namely:

$$\begin{aligned} P_{k0}z_{k+2} - (P_{k1} + P_{k0} + P_{k-1,0})z_{k+1} + (P_{k1} + P_{k2} + 2P_{k-1,0} + \beta)z_k \\ -(P_{k2} + P_{k3} + P_{k-1,0})z_{k-1} + P_{k3}z_{k-2} = \beta y_k \end{aligned}$$

where

$$\begin{aligned} P_{k0} &= \frac{\alpha_{k+1}}{(x_{k+2} - x_{k+1})(x_{k+1} - x_k)} \\ P_{k1} &= \frac{\alpha_{k+1} + \alpha_k}{(x_{k+1} - x_k)^2} \\ P_{k2} &= \frac{\alpha_{k-1} + \alpha_k}{(x_k - x_{k-1})^2} \\ P_{k3} &= \frac{\alpha_{k-1}}{(x_k - x_{k-1})(x_{k-1} - x_{k-2})} \end{aligned}$$

and $z_{-2} = z_{-1} = z_{n+1} = z_{n+2} = 0$.

The parameter β is chosen such that it satisfies:

$$\beta \gg \frac{1}{\min_k (x_{k+1} - x_k)^2}$$

which implies diagonal dominance for the system of equations (10.23).

To find discontinuities in the function or for its derivative we look at zero crossings of the error between the observation and the desired data

$$z_i - y_i \tag{10.23}$$

and at zero crossings of the approximate curvature

$$\frac{z_{i+1} - z_i}{x_{i+1} - x_i} - \frac{z_i - z_{i-1}}{x_i - x_{i-1}} \tag{10.24}$$

Slope discontinuities are characterized by successive zero crossings of type (10.23) and function discontinuities are characterized by zero crossings of type (10.23) and (10.24).

For our problem we are interested in eliminating only the points which are associated with the shocks. This is a trade-off between obtaining a *smoother* function and preserving as much as possible the discontinuous character of the problem for a given time interval. If one would like to single out a region of the solution (among the five regions of the flow described earlier) with the greatest influence during numerical optimization, one would select the points where the shock occurs.

For this reason we restricted the algorithm of discontinuity detection to eliminate only the shock points.

The detection of shock points was performed by considering only points with approximate curvatures above a certain threshold value. This approach was suggested to us by the fact that the curvature for the analytical solution is very steep in the shock region.

The result of discontinuities detection is shown in Fig. 10.8.

10.5 Overview of numerical results

Our goal was to control the location of the discontinuities by matching the numerical flow to observations that contain the *desired* location of discontinuities.

For many problems (including ours) the problem of finding a "matching" flow at a given time is equivalent to the problem of finding the corresponding vector of initial conditions (the initial conditions serving as the control variables in the optimal control setting).

For practical applications it is more important to consider the impact of the change of shock location on the flow parameters rather than the explicit description of the *new* discontinuity location. For this reason we concentrated our research efforts on matching the flow to a *desired* flow rather than introducing the explicit shock location as a variable in the optimal control setup (as performed by Cliff et al. [36] for duct flow with quasi 1-D Euler equations).

We used the discrete forward model to obtain the tangent linear model and then the adjoint model, which provides the gradient or a subgradient of the cost functional to the smooth (non smooth) minimizer.

If the location of the discontinuities were to be introduced as an explicit variable, then the original model should be modified to accommodate the new requirements, a change requiring complex adjustments. This is one of the arguments supporting our claim that our approach is more appropriate for practical optimization problems involving discontinuities.

We considered as forward models the artificial viscosity model **AVM** and the high-resolution model **HRM**. For each of the two numerical models we employed unconstrained optimization methods (**L-BFGS** algorithm for smooth optimization and **PVAR** algorithm for non smooth optimization) described in chapter 4.

The control variables were chosen to be the initial parameters to the left and to the right of the membrane: pressure p_L, p_R and density ρ_L, ρ_R.

The desired observations were obtained as exact solutions of the shock-tube problem at times $t = 0.15$ or $t = 0.24$, starting with prescribed initial conditions.

We considered three sets of initial parameters, which are referred to as:

- the first set of parameters ($\mathcal{FSP}$)
$$\mathcal{FSP} = [\rho_L = 1.1, p_L = 1.1, \rho_R = 0.2, p_R = 0.2]$$
- the second set of parameters ($\mathcal{SSP}$)
$$\mathcal{SSP} = [\rho_L = 1.2, p_L = 1.2, \rho_R = 0.3, p_R = 0.3]$$
and the third set of parameters ($\mathcal{TSP}$):
$$\mathcal{TSP} = [\rho_L = 2.5, p_L = 2.0, \rho_R = 0.5, p_R = 0.6]$$

The initial guess for both minimization methods ($\mathcal{INIT}$) is characteristic for the Sod shock-tube problem ([177]):

$$\mathcal{INIT} = [\rho_L = 1.0, p_L = 1.0, \rho_R = 0.1, p_R = 0.1]$$

The initial values for velocities to the left and to the right of the membrane were taken to be zero.

The flow obtained using the first set ($\mathcal{FSP}$), the second set ($\mathcal{SSP}$) and the third set of parameters ($\mathcal{TSP}$) as initial conditions is compared in Fig. 10.1, 10.2 and 10.3 to the flow obtained using the initial guess ($\mathcal{INIT}$) as initial conditions. It can be seen, especially from the plot corresponding to the third set of parameters ($\mathcal{TSP}$), that there is a large discrepancy between the initial guess and the observations. Despite this discrepancy the minimization will be performed succesfully (using the nonsmooth minimizer **PVAR**) and its numerical results are shown in Fig. 10.19, 10.20 and 10.18.

The numerical results for both models (**AVM** and **HRM**), using each of the optimization methods (**L-BFGS** and **PVAR**), are presented in Fig. 10.4 - 10.5 (the evolution of the cost functional vs. the number of minimization iterations) and in Fig. 10.6 - 10.7 (the numerical flow obtained using the results of optimization compared with the observations). The values of the optimized control parameters are presented in Table 10.1 (**HRM**) and Table 10.2 (**AVM**).

We considered two different time horizons for the optimal control problem: $T_W = 0.15$ or $T_W = 0.24$ (in non dimensional units). They were chosen for two main reasons. First, at the end of the time window the flow exhibits all five regions of discontinuities previously discussed. Second, if one increases the time horizon from $T_W = 0.24$ to a slightly larger value $time = 0.3$, one can see from Fig. 10.9 that several characteristics of the discontinuities have already disappeared. from the spatial domain considered.

We also employed two expressions for the cost functional, with observations located either at the end of the assimilation window or with distributed observations in time.

When the observations were located at the end of the time window ($t = T_W$) the following discrete form of the cost functional was considered:

$$\mathcal{J}(\mathbf{U}(\cdot,0),\mathbf{P}(\cdot,0),\rho(\cdot,0)) = \sum_{i=1}^{Npoints}\Big(W_{\mathbf{U}}(i)\times(\mathbf{U}^{num}(i)-\mathbf{U}^{obs}(i))^2$$
$$+W_{\mathbf{P}}(i)\times(\mathbf{P}^{num}(i)-\mathbf{P}^{obs}(i))^2+W_\rho\times(\rho^{num}(i)-\rho^{obs}(i))^2\Big)$$

where

$$\mathbf{U}(x,0)=\begin{cases}0.0, & x<0.5\\ 0.0, & x>0.5\end{cases}\quad \mathbf{P}(x,0)=\begin{cases}p_L, & x<0.5\\ p_R, & x>0.5\end{cases}\quad \rho(x,0)=\begin{cases}\rho_L, & x<0.5\\ \rho_R, & x>0.5\end{cases}$$

with $(\rho_L, p_L, \rho_R, p_R)$ the control variables described above.

$Npoints$ is the number of points for space discretization, $W_{\mathbf{U}}, W_{\mathbf{P}}, W_\rho$ are the weights matrices attached to points (we considered $weight = 0.0$, $weight = 1.0$ or $weight = 25.0$), $\mathbf{U}^{num}, \mathbf{P}^{num}, \rho^{num}$ are the fields of velocity, pressure and density at time t_{final} while $\mathbf{U}^{obs}, \mathbf{P}^{obs}, \rho^{obs}$ are the observations for velocity, pressure and density.

For distributed observations the discrete form of the cost functional is:

$$\boldsymbol{\mathcal{J}}(\mathbf{U}(\cdot,0),\mathbf{P}(\cdot,0),\rho(\cdot,0)) = \sum_{j=1}^{Nobs}\sum_{i=1}^{Npoints}\Big(W_{\mathbf{U}}(i)\times\left(\mathbf{U}^{num}_{(j)}(i)-\mathbf{U}^{obs}_{(j)}(i)\right)^2$$
$$+W_{\mathbf{P}}(i)\times\left(\mathbf{P}^{num}_{(j)}(i)-\mathbf{P}^{obs}_{(j)}(i)\right)^2+W_\rho\times\left(\rho^{num}_{(j)}(i)-\rho^{obs}_{(j)}(i)\right)^2\Big)$$

In addition to the notations for the previous cost functional we denote by $Nobs$ the number of instances during the assimilation window for which we consider the observations, $\mathbf{U}^{num}_{(j)}, \mathbf{P}^{num}_{(j)}, \rho^{num}_{(j)}$ are the fields of velocity, pressure and density at time $t_{(j)}$, ($1 \le j \le Nobs$) while $\mathbf{U}^{obs}_{(j)}, \mathbf{P}^{obs}_{(j)}, \rho^{obs}_{(j)}$ are the observations for velocity, pressure and density at the same observation times $t_{(j)}$, respectively.

10.6 Numerical results for the high-resolution model HRM

For the **HRM** model the optimized values of the control parameters are in excellent agreement with the parameters' *desired* values for both assimilation windows when the non smooth optimization package **PVAR** was employed. Fig. 10.4 shows a decrease of more than 2 orders of magnitude for the cost functional. The *optimized* values of the control parameters $[\rho_L, p_L, \rho_R, p_R]$ obtained as a result of non smooth minimization (the row **PVAR** in Table 10.1) display a very good agreement with the *desired* parameters. This remark is also supported by Fig. 10.6 which presents the comparison between the numerical *optimized* solution and the observations.

For the model **HRM** we also employed a cost functional with time-distributed observations for the larger time window $T_W = 0.24$. The optimized values of the control parameters obtained as a result of the non smooth minimization are shown as entries in the column **PVAR** [*d.c.*] (distributed controls) in Table 10.1. Since we have already obtained excellent *optimized* results (almost identical to the *desired* values of the parameters) for a cost functional computed using only final time observations, we may conclude that the additional information provided by time-distributed observations was extraneous.

To verify the robustness of our approach we considered the third set of parameters ($\mathcal{TSP}$). The results obtained using non smooth optimization **PVAR** (also shown in Table 10.1) are in very good agreement with the *desired* values of the parameters. Fig. 10.10 shows that the flow obtained with the *optimized* control parameters as initial conditions matches closely the observations. We can also see that the new location of discontinuities matches the *desired* location.

The evolution of the numerical optimal solution obtained during different stages of optimization versus the observations was plotted in Fig. 10.19, 10.20 and 10.18. It can be seen from these plots that the numerical solution obtained during minimization has the characteristics of a solution acceptable from the physical point of view. Another argument in favor of this affirmation is presented next, using the physical entropy.

The evolution of the entropy during various stages of the minimization process (computed at the end of the assimilation window) is displayed in Fig. 10.11 and Fig. 10.12. It is known that the correct weak solution should satisfy the entropy condition, which states that the entropy of fluid particles does not decrease. Over the contact discontinuity the entropy decreases, but since fluid particles do not cross the contact discontinuity, the entropy of the particles does not decrease. This shows that the numerical solution has indeed the characteristics of a physical solution.

The **L-BFGS** minimization converged to the *desired* parameters only for the shorter time window ($T_W = 0.15$). For the larger time window ($T_W = 0.24$) the **L-BFGS** minimization failed.

10.7 Numerical results for the artificial viscosity model AVM

We also applied successfully the non smooth minimization algorithm **PVAR** to the artificial viscosity model **AVM** which represents the class of models which *smooth* the discontinuities (see Table 10.2).

As seen in Table 10.2, **L-BFGS** converged only for the time window $T_W = 0.15$, although the cost functional becomes "numerically" *smoother* for **AVM** model. For the larger time window $T_W = 0.24$ **L-BFGS** proved useful in a different setting. By using the **L-BFGS** output as an initial guess for the **PVAR** method we obtained convergence to the *desired* parameters in fewer minimization iterations. The values of the control parameters obtained using this approach are shown in Table 10.2 (row **PVAR** [*input*]).

Scaling for the gradient of the cost functional was applied to the cases when **L-BFGS** unconstrained optimization failed. The scaling was chosen such that all components of the gradient have numerical values of order one. The scaled gradient **L-BFGS** optimization did not converge to the *desired* values of control parameters (rows **L-BFGS** [*s.*] (scaled) of Table 10.2).

To alleviate the impact of discontinuities we tested a method whereby we selectively applied weights to the points of discontinuities. The choice of the points where weights are applied was based on the trade-off between the desire for a *smoother* function and the requirement of preserving as much as possible of the properties for the original problem. Thus we assigned weights only to points where the shock occurred and we did not consider contact discontinuities nor rarefaction waves. Fig. 10.8 shows the "shock" points after they were selected using the discontinuity detection method described earlier in this chapter.

Different weights were considered in the computation the cost functional and its gradient ($weight = 0.0$ corresponds to removal of these points from the cost functional and its gradient computation, $weight = 1.0$ means that all the mesh points are considered to have the same influence while for $weight = 25.0$ the influence of the shock is dominant). Since the shock location changes in the forward model after each minimization iteration the method of discontinuity detection was re-applied and corresponding shock points were found.

The weighted minimization with $weight = 25.0$ failed for both time windows $T_W = 0.15$ and $T_W = 0.24$.

For $weight = 0.0$ a successful minimization was obtained (see row **L-BFGS** [$w = 0$] of Table 10.2). The values of the control vector obtained using this approach were similar in quality to the values obtained with **PVAR** and no weight considerations.

10.8 Additional numerical considerations

The control variables employed for this research were the initial values for pressure and density. Since the desired value for the initial velocity is 0.0 (both to the left and to the right of the membrane) we did not consider the initial value of the velocity among the control variables. Another reason for selecting the initial values of the velocity to be zero is related to the physical aspects of the shock-tube problem. If the initial values for the velocity are considered as control variables, then during the minimization their updates may have values which are not physical or the corresponding adjoint variables may lead to solutions developing bifurcation points (Cacuci [26]).

Comparing the three sets of *desired* parameters one may argue that the distance between the first or second set of parameters ($\mathcal{FSP}$) or ($\mathcal{SSP}$) and the initial guess ($\mathcal{INIT}$) is rather small. But comparing the flow corresponding to ($\mathcal{INIT}$) with the flow obtained for either ($\mathcal{FSP}$) or ($\mathcal{SSP}$) (Fig. 10.1 and 10.2) one notices large differences in the location of discontinuities and the values for the flow variables, which provides a very good argument for our choices. The third set of parameters ($\mathcal{TSP}$) was chosen to be at a much larger distance to the initial guess ($\mathcal{INIT}$) in order to test the robustness of our approach.

We would like to describe in more detail the " failure" of the **L-BFGS** method for our problem. For some cases (e.g., for **HRM** model using the first set of observations and time window $T_W = 0.24$) the minimization *per se* performed successfully from the optimization point of view (i.e., decrease of the cost functional and update of the vector of control variables using a computed new step size). The failure is due to the fact that the updated vector of control variables did not qualify as a solution from the physical point of view.

Although the non smooth minimization algorithm **PVAR** performed successfully for both numerical models there are large differences in the memory and CPU time requirements. For comparable accuracy the number of mesh points for **HRM** model was 200 while it was 500 for **AVM**, with corresponding differences in the number of time steps required.

The influence of the numerical model over the optimization results for the first two sets of observations is presented in Fig. 10.14 and 10.15 for the non smooth optimization algorithm **PVAR**, respectively in Fig. 10.17 and 10.16 for the **L-BFGS** minimization algorithm.

Table 10.1. Optimization results for the high-resolution model

FIRST SET OF OBSERVATIONS AND TIME=0.15				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	1.1	1.1	0.2	0.2
L-BFGS	1.10143	1.10251	0.19934	0.19865
PVAR	1.10059	1.10187	0.19942	0.19884

FIRST SET OF OBSERVATIONS AND TIME=0.24				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	1.1	1.1	0.2	0.2
L-BFGS	Failed	Failed	Failed	Failed
PVAR	1.09815	1.08966	0.19993	0.19894
L-BFGS [s.]	1.04032	0.99664	0.13887	0.99628
PVAR [d.c.]	1.10088	1.10915	0.20122	0.19886

SECOND SET OF OBSERVATIONS AND TIME=0.15				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	1.2	1.2	0.3	0.3
L-BFGS	1.20161	1.20342	0.29712	0.29973
PVAR	1.20052	1.20278	0.29752	0.29953

SECOND SET OF OBSERVATIONS AND TIME=0.24				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	1.2	1.2	0.3	0.3
L-BFGS	1.03479	0.85757	0.35072	0.25325
PVAR	1.19406	1.19203	0.30308	0.29946
L-BFGS [s.]	1.38023	0.84461	0.37357	0.26728
PVAR [d.c.]	1.20689	1.20698	0.30294	0.29962

THIRD SET OF OBSERVATIONS AND TIME=0.24				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	2.5	2.0	0.5	0.6
PVAR	2.49591	1.97919	0.49941	0.60096

Table 10.2. Optimization results for the artificial viscosity model

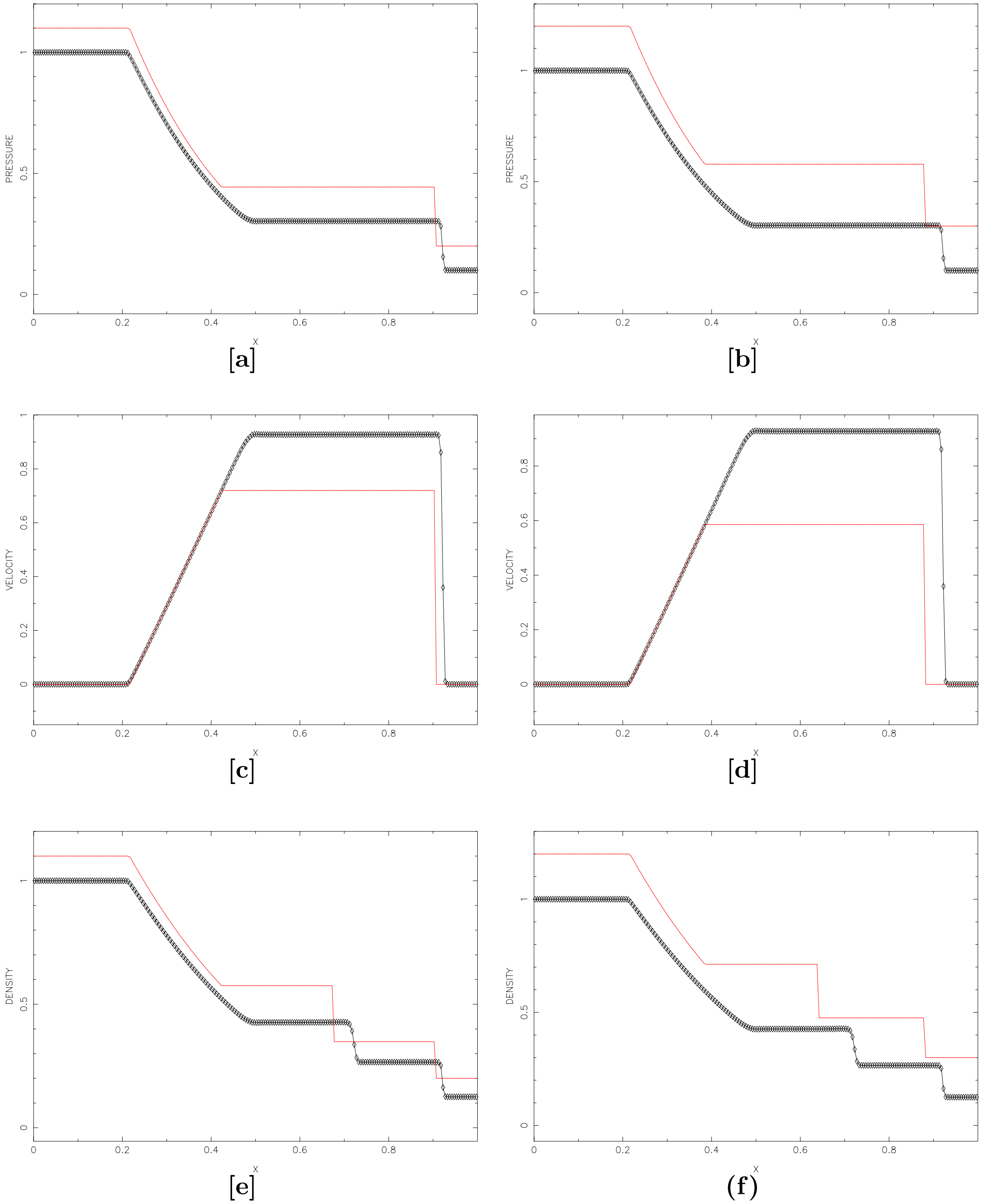
FIRST SET OF OBSERVATIONS AND TIME=0.15				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	1.1	1.1	0.2	0.2
L-BFGS	1.09712	1.09947	0.20432	0.19756
L-BFGS [w=0]	1.10031	1.10459	0.20514	0.19786
PVAR	1.09685	1.09933	0.20439	0.19782

FIRST SET OF OBSERVATIONS AND TIME=0.24				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	1.1	1.1	0.2	0.2
L-BFGS	1.02638	1.00347	0.18012	0.19296
L-BFGS [w=0]	1.09742	1.10173	0.20004	0.20154
PVAR	1.09737	1.09966	0.20357	0.19874
PVAR [input]	1.09741	1.09961	0.20344	0.19867
L-BFGS [s.]	1.03685	0.96042	0.13276	0.35517

SECOND SET OF OBSERVATIONS AND TIME=0.15				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	1.2	1.2	0.3	0.3
L-BFGS	1.19784	1.19856	0.30582	0.29714
L-BFGS [w=0]	1.19327	1.18962	0.29983	0.30134
PVAR	1.19768	1.19778	0.30583	0.29702

SECOND SET OF OBSERVATIONS AND TIME=0.24				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	1.2	1.2	0.3	0.3
L-BFGS	1.19832	1.19846	0.30615	0.29891
L-BFGS [w=0]	1.19872	1.19641	0.30373	0.29778
PVAR	1.19764	1.19825	0.30527	0.29735

THIRD SET OF OBSERVATIONS AND TIME=0.24				
Parameter	ρ_L	p_L	ρ_R	p_R
Desired	2.5	2.0	0.5	0.6
L-BFGS [s.]	2.488975	1.904185	0.65136	0.83691



[a] [b] [c] [d] [e] (f)

Figure 10.1. Pressure, velocity and density: initial guess (◇) and exact observation (red line) at time=0.24 for the **HRM** model: first set of observations ([**a**], [**c**], [**e**]) and the second set of observations ([**b**], [**d**], [**f**])

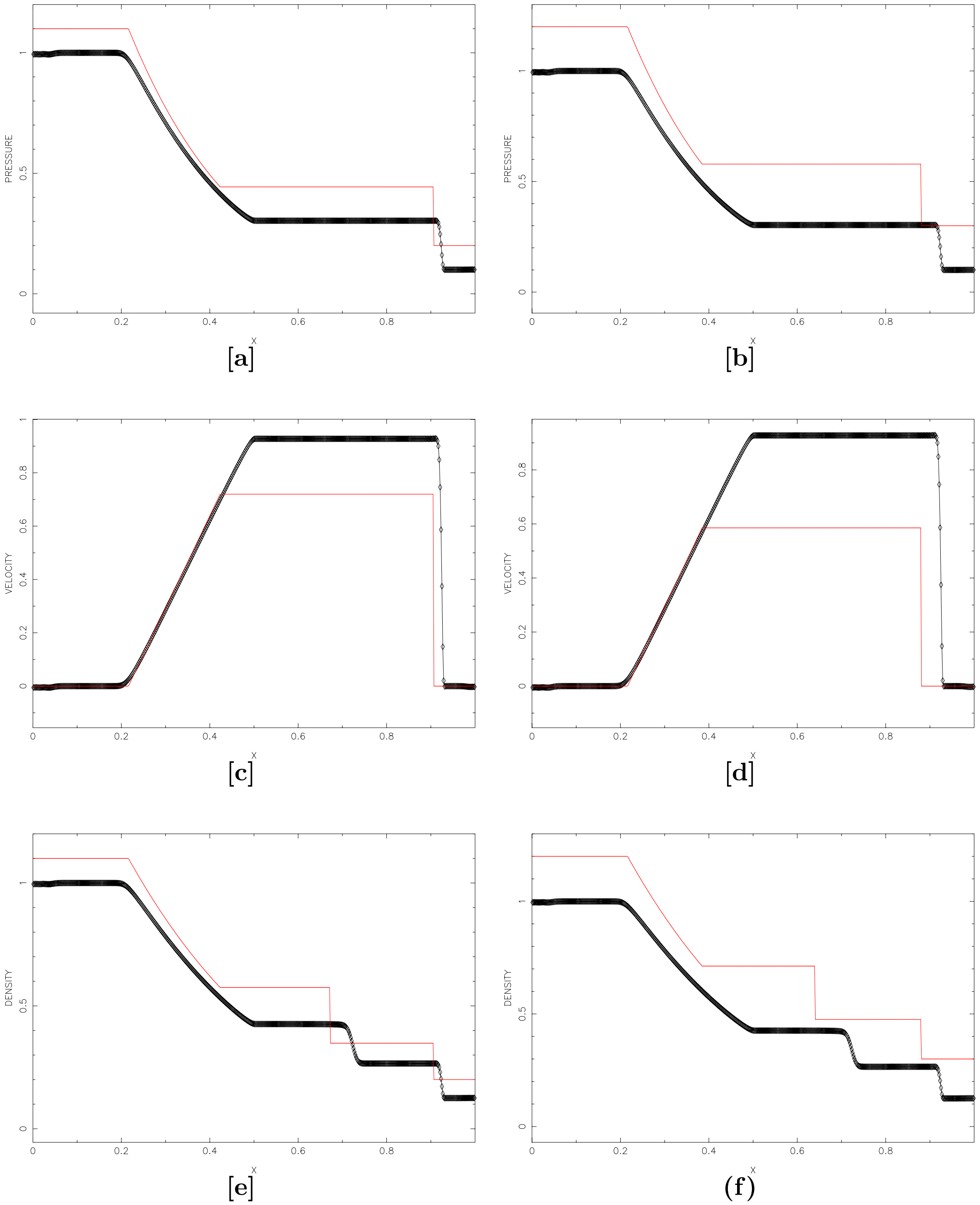
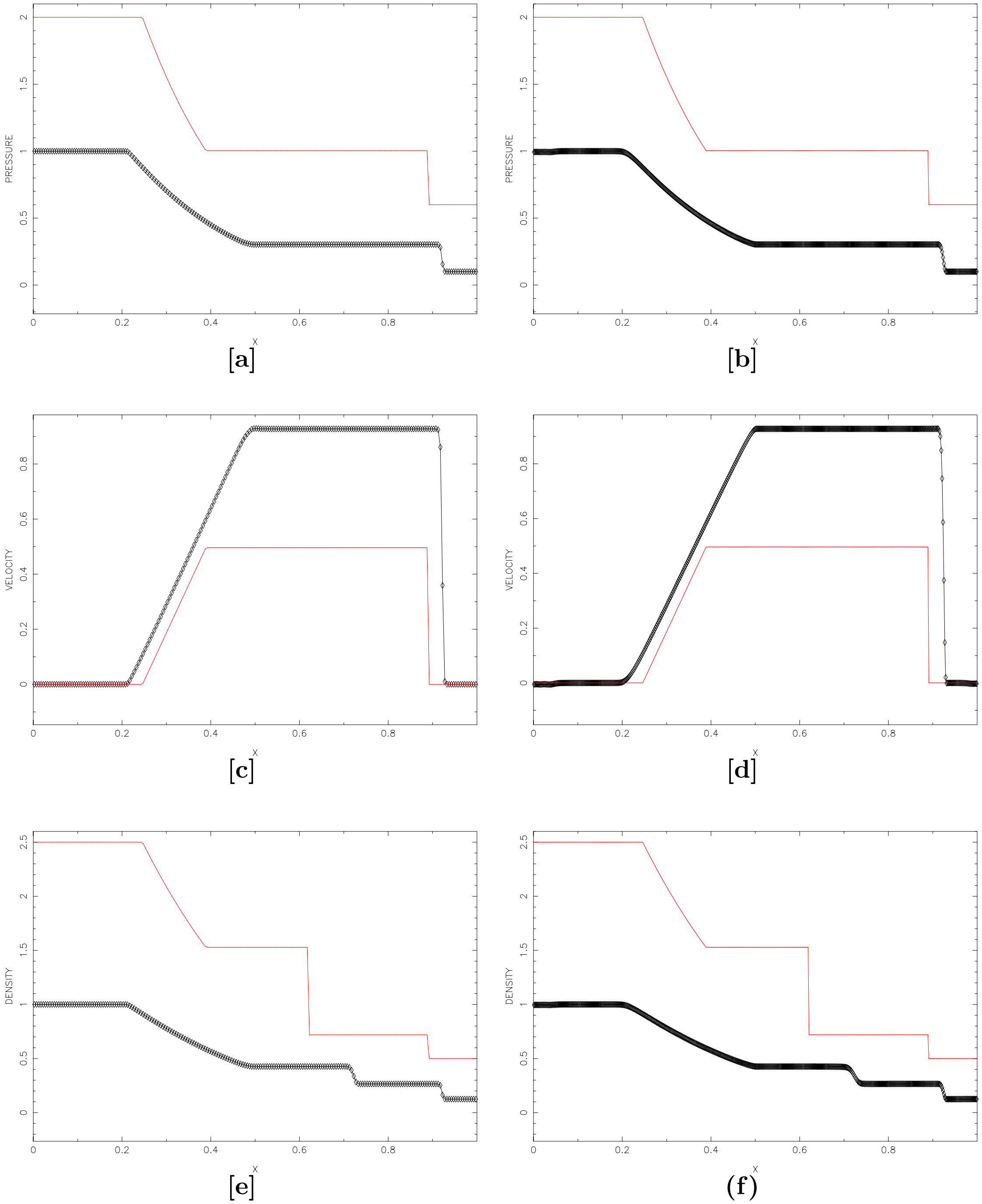


Figure 10.2. Pressure, velocity and density: initial guess (⋄) and exact observation (red line) at time=0.24 for the **AVM** model: first set of observations ([**a**], [**c**], [**e**]) and the second set of observations ([**b**], [**d**], [**f**])



[a] [b]

[c] [d]

[e] (f)

Figure 10.3. Pressure, velocity and density: initial guess (⋄) and exact observation (red line) at time=0.24 for the third set of observations: for the **HRM** model ([**a**], [**c**], [**e**]) and for the **AVM** model ([**b**], [**d**], [**f**])

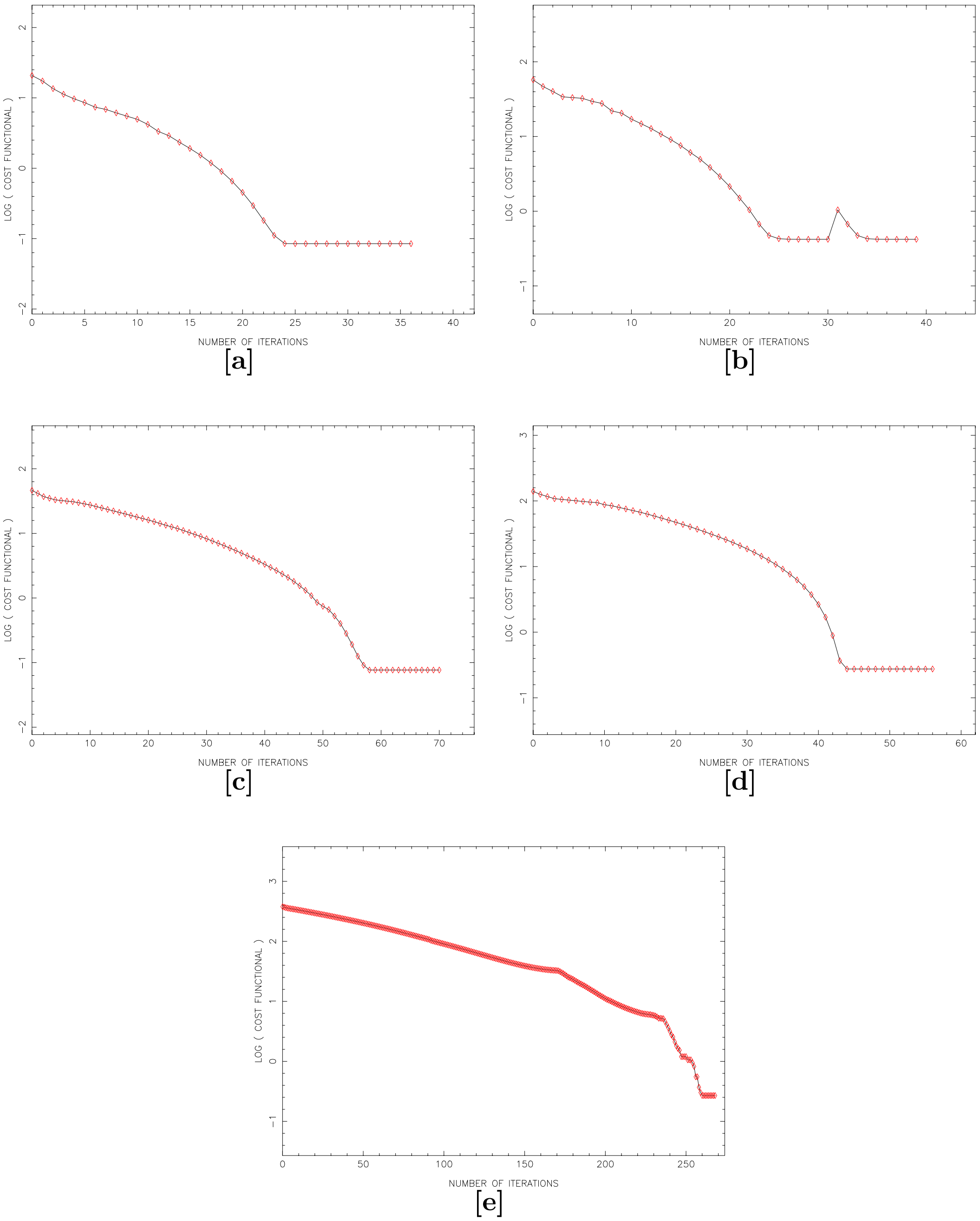
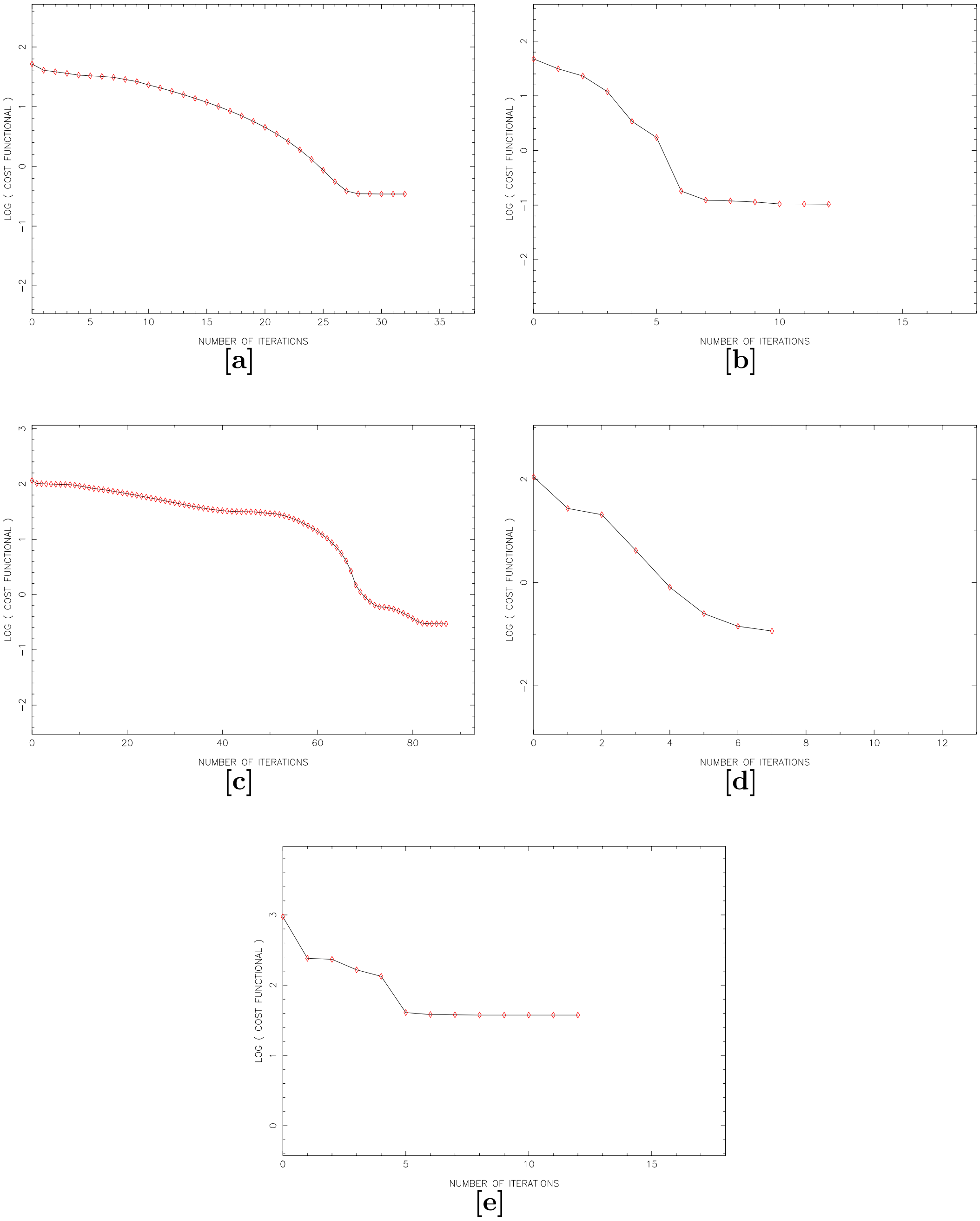


Figure 10.4. Evolution of the logarithm of cost functional vs. number of iterations during non smooth minimization **PVAR** for the **HRM** model at time=0.24: first set of observations without ([**a**]) or with ([**b**]) distributed observations; second set of observations without ([**c**]) or with ([**d**]) distributed observations; third set of observations [**e**]



[a]

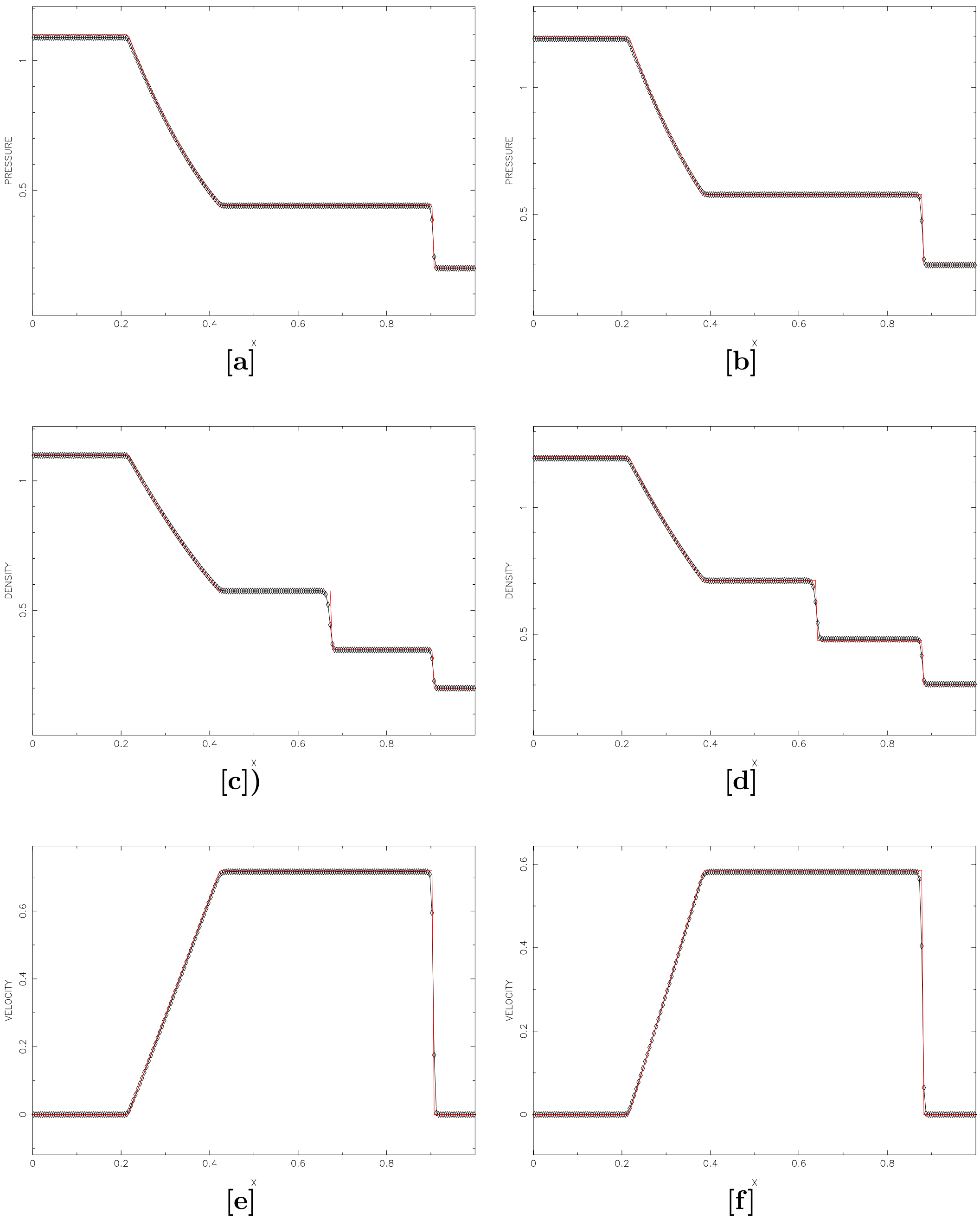
[b]

[c]

[d]

[e]

Figure 10.5. Evolution of the logarithm of cost functional vs. number of minimization iterations with the **AVM** model at time=0.24: non smooth optimization **PVAR** and first ([**a**]) or second ([**c**]) set of observations; **L-BFGS** optimization with weight=0.0 for the first ([**b**]) or second ([**d**]) set of observations; (**L-BFGS**) scaled optimization for the final set of observations [**e**]



[**a**] [**b**]

[**c**]) [**d**]

[**e**] [**f**]

Figure 10.6. Pressure, density and velocity: observations (red line) and numerical solution ($\diamond$) of non smooth optimization **PVAR** for the **HRM** model at time=0.24: first set of observations ([**a**], [**c**], [**e**]) and second set of observations ([**b**], [**d**] and [**f**])

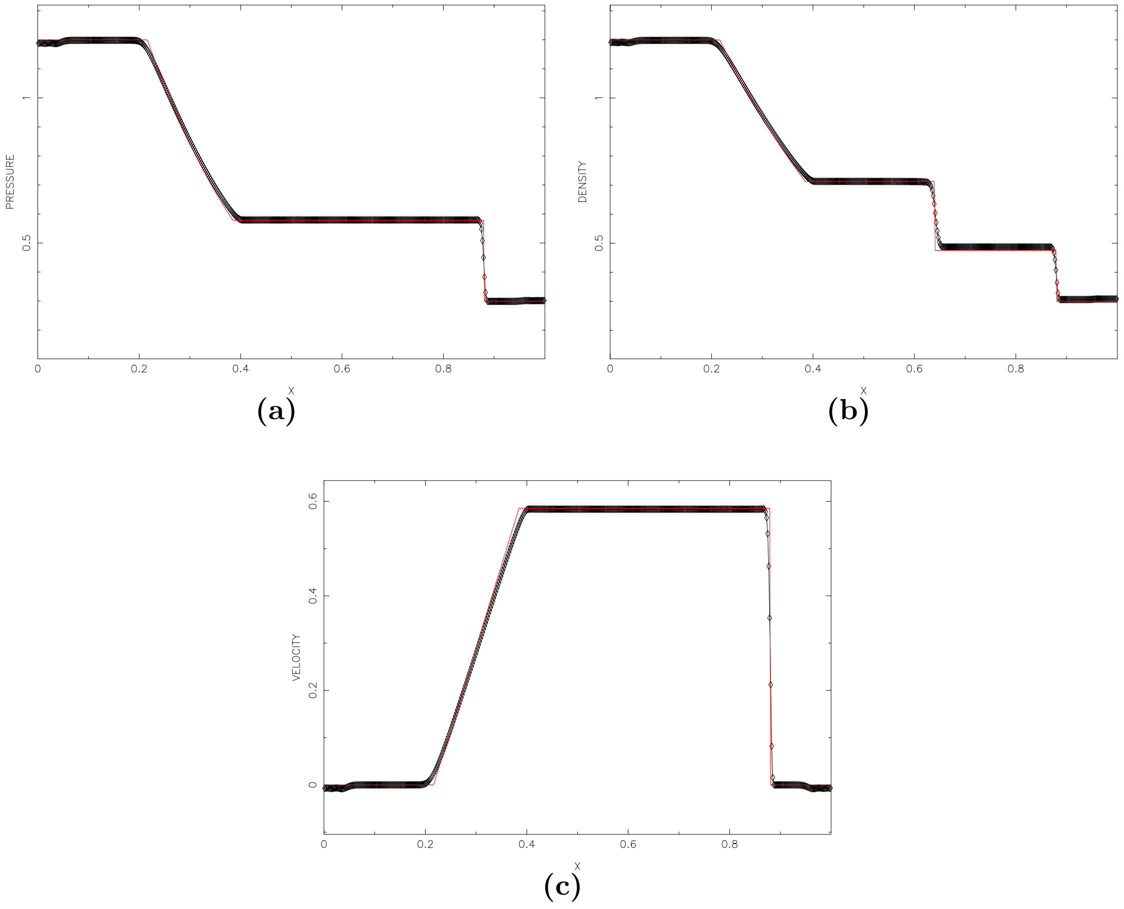
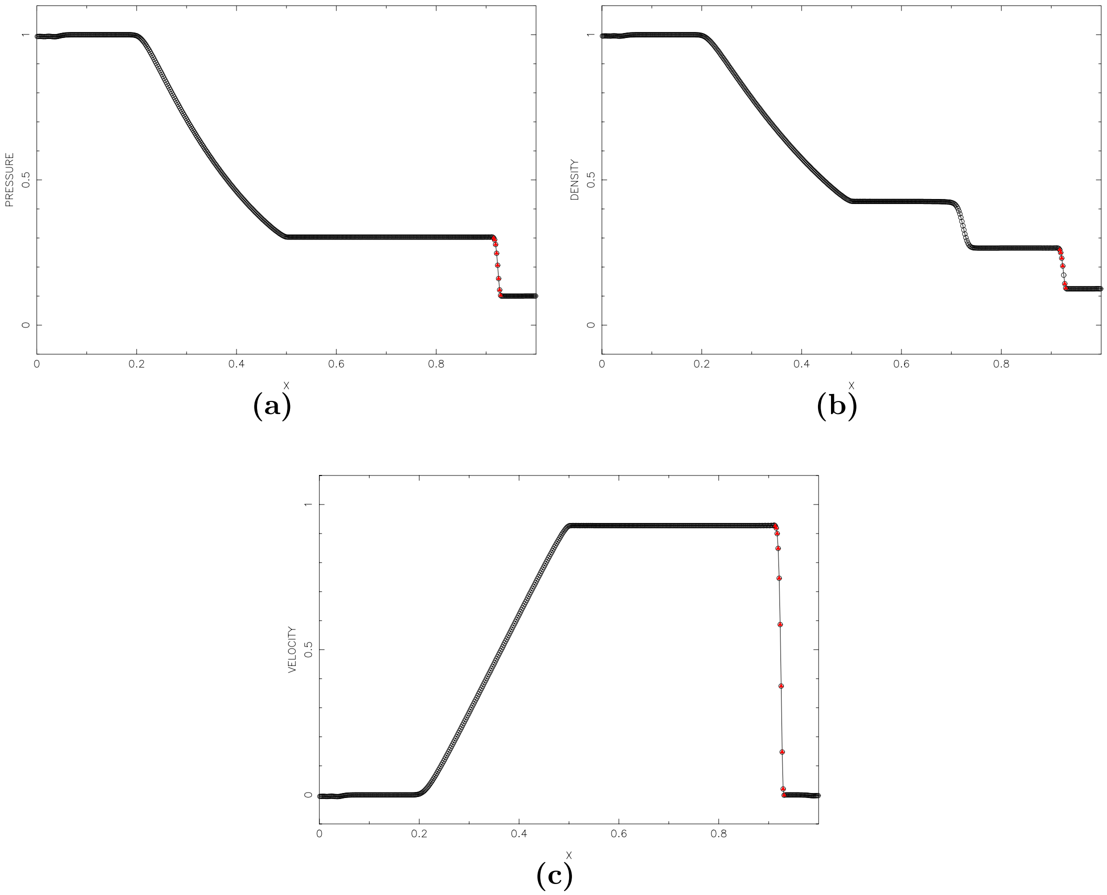


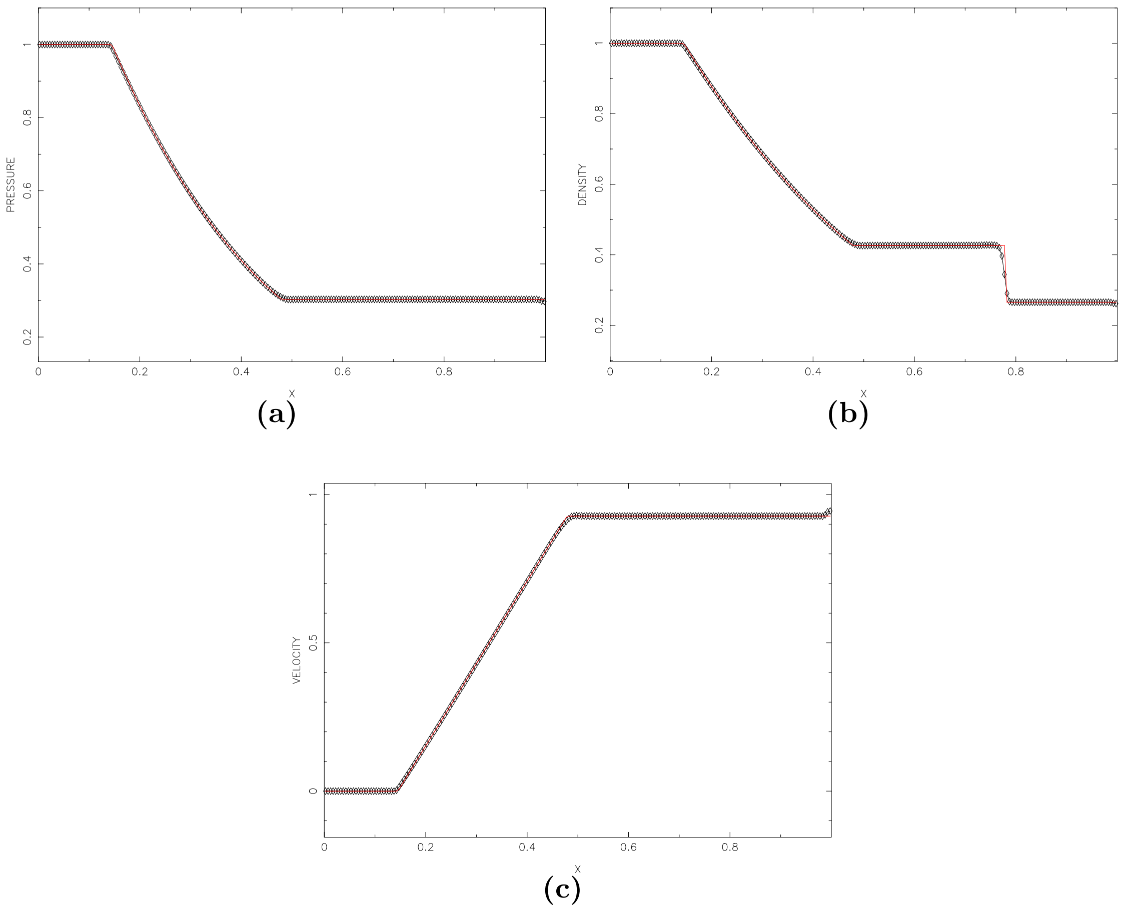
Figure 10.7. Pressure ([**a**]), density ([**b**]) and velocity ([**c**]): first set of observations (red line) and numerical solution ($\diamond$) of **L-BFGS** optimization (with weight=0.0) for the **AVM** model at time=0.24.



(a) **(b)**

(c)

Figure 10.8. Discontinuity detection for the **AVM** model: the selected points (red) for pressure ([**a**]), density ([**b**]) and velocity ([**c**])

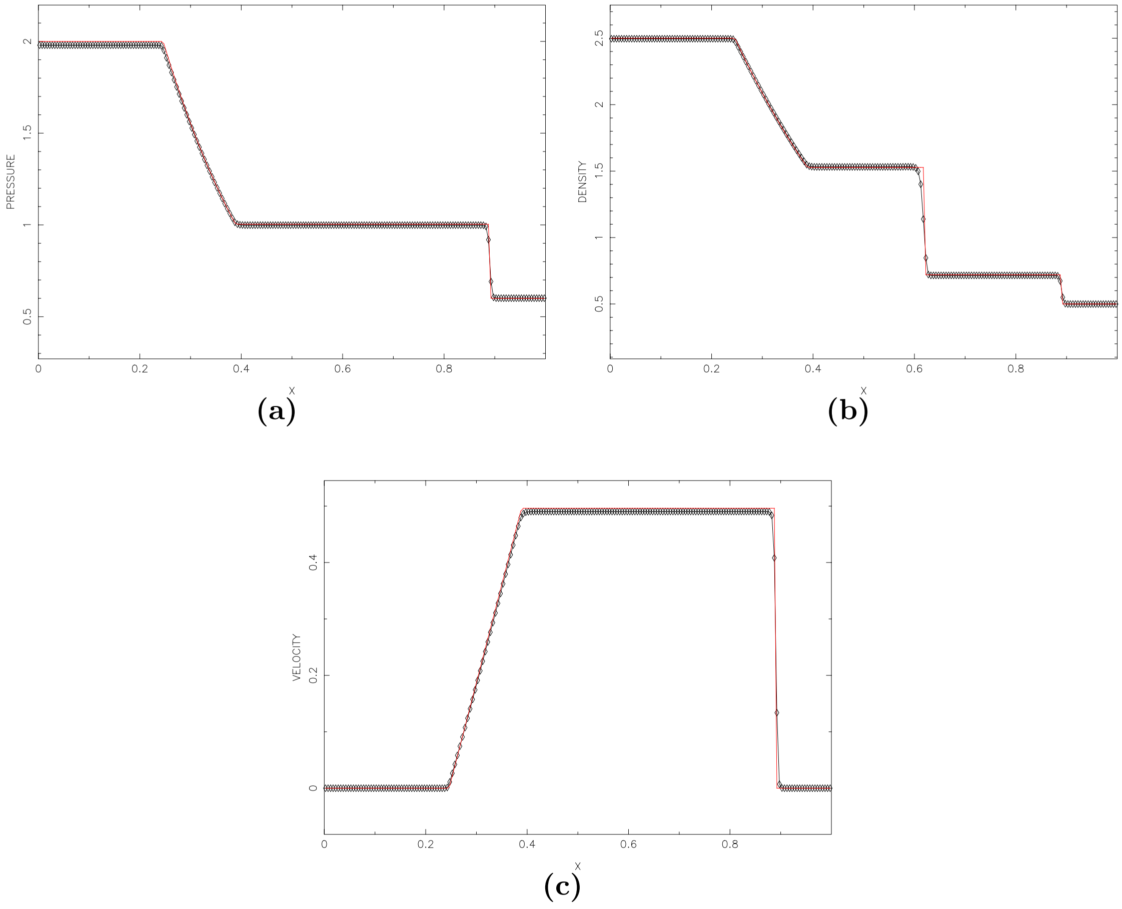


(a)

(b)

(c)

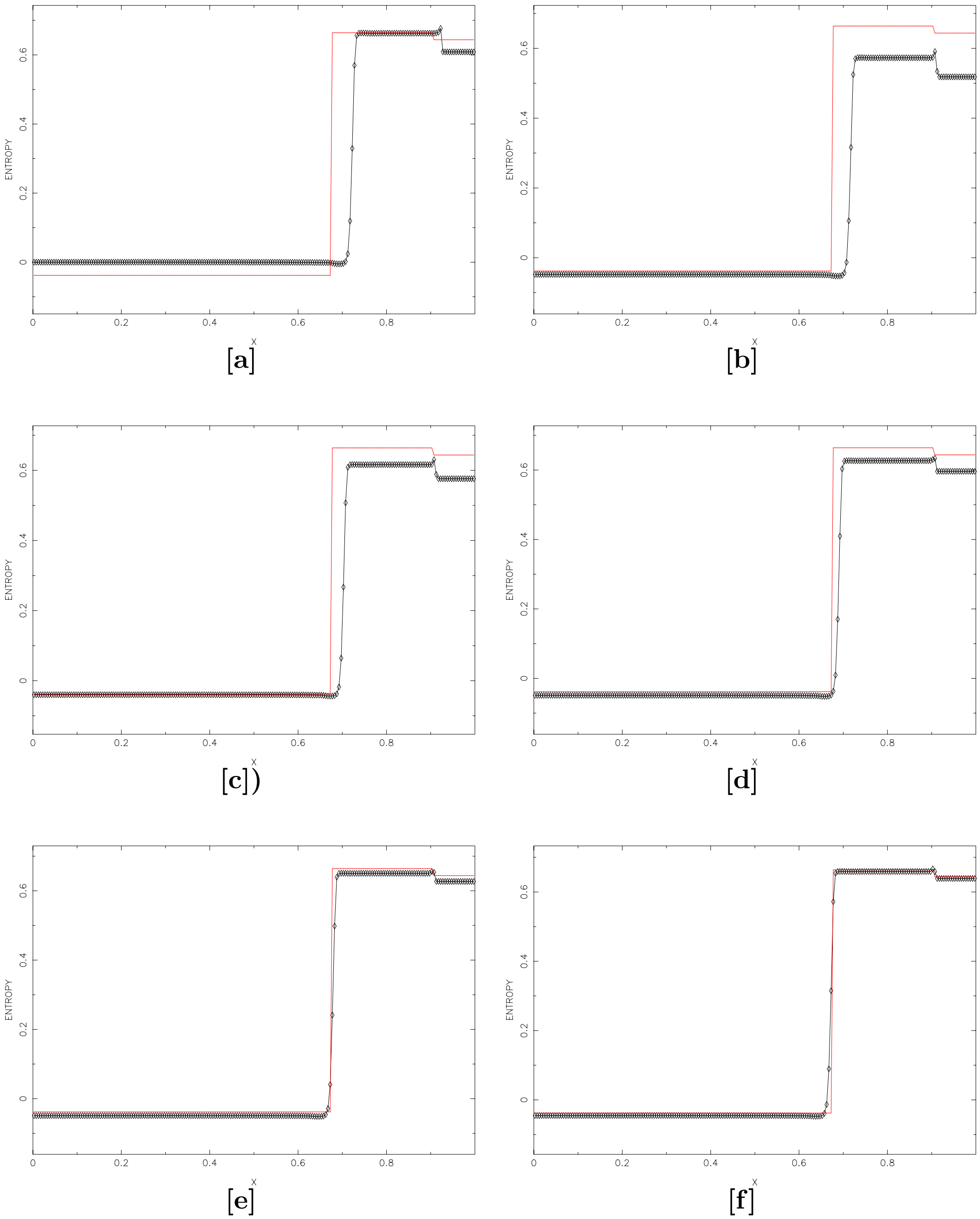
Figure 10.9. Pressure ([**a**]), density ([**b**]) and velocity ([**c**]): Numerical ($\diamond$) and analytical (red line) solution of high-resolution model for the shock-tube problem at time=0.30



(a) **(b)**

(c)

Figure 10.10. Pressure ([**a**]), density ([**b**]) and velocity ([**c**]): observations (red line) and numerical solution($\diamond$) of **PVAR** for the **HRM** model at time=0.24 for the final set of observations

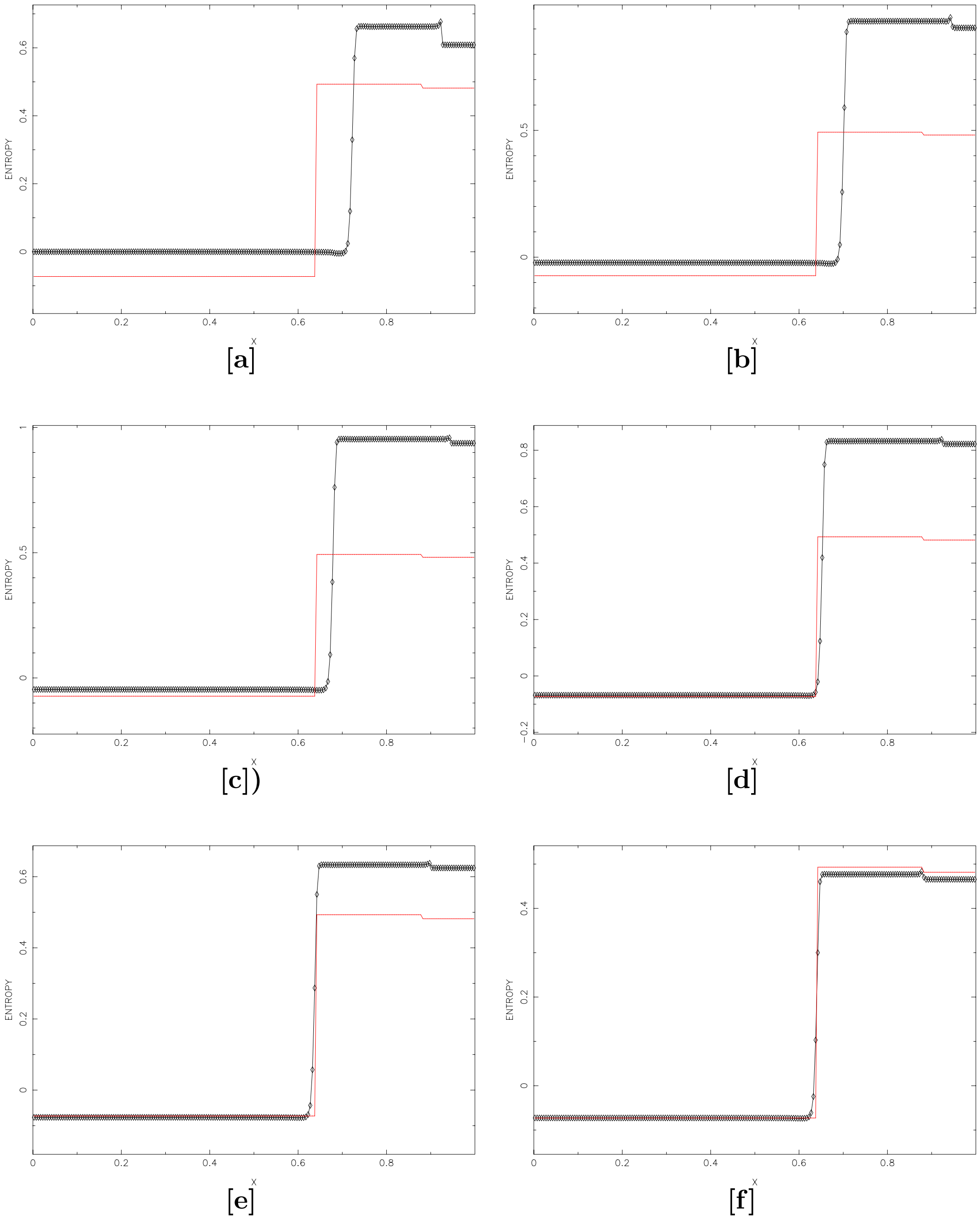


[**a**] [**b**]

[**c**]) [**d**]

[**e**] [**f**]

Figure 10.11. Evolution of numerical ($\diamond$) and analytical (red line) entropy (shown at final time $t = 0.24$ for the **HRM** model and for the first set of observations) during non smooth minimization **PVAR**: [**a**] iteration=0, [**b**] iteration=5, [**c**] iteration=10, [**d**] iteration=15, [**e**] iteration=20, [**f**] final iteration

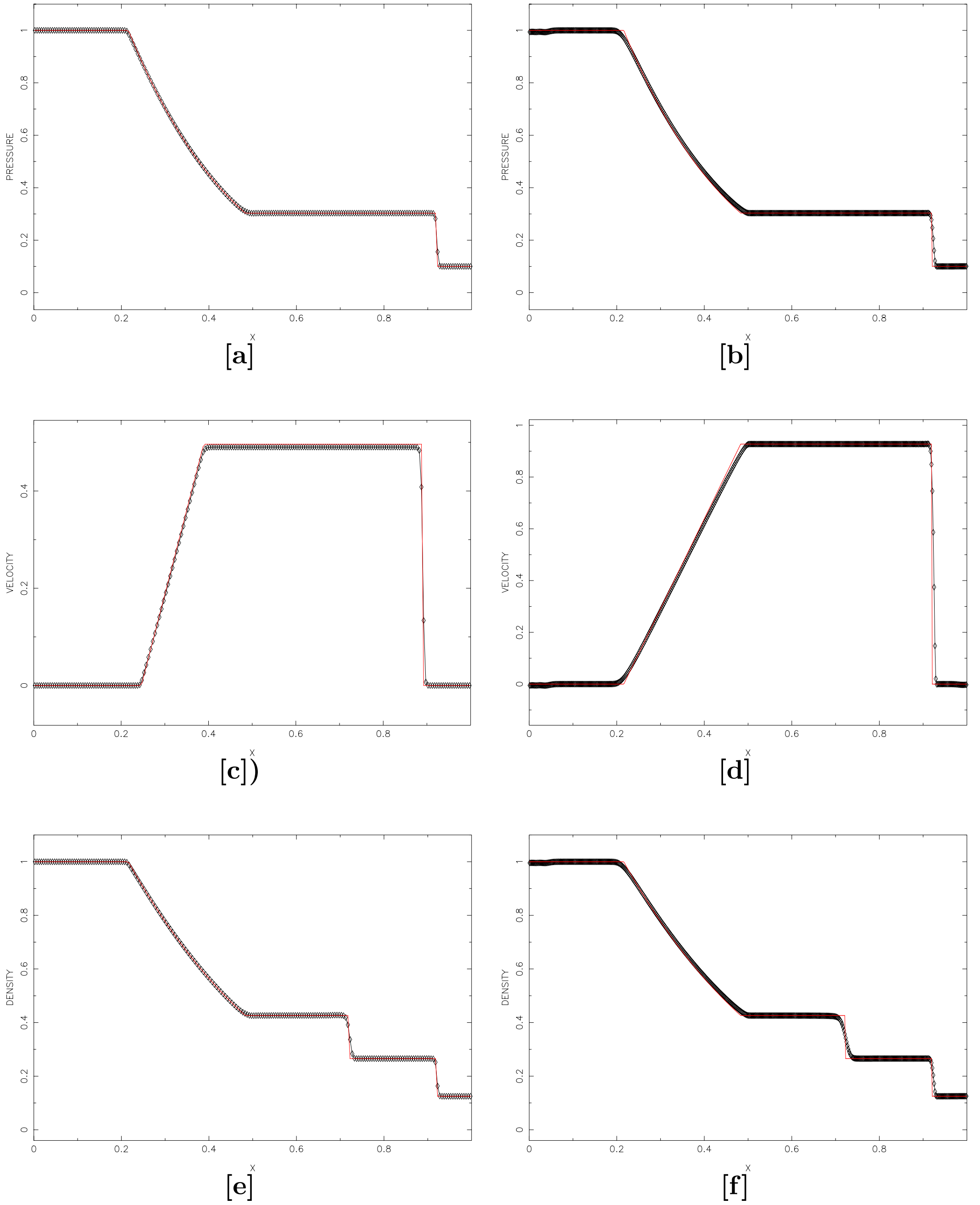


[**a**] [**b**]

[**c**]) [**d**]

[**e**] [**f**]

Figure 10.12. Evolution of numerical (⋄) and analytical (red line) entropy (shown at final time $t = 0.24$ for the **HRM** model and for the second set of observations) during non smooth minimization **PVAR**: [**a**] iteration=0, [**b**] iteration=10, [**c**] iteration=20, [**d**] iteration=35, [**e**] iteration=45, [**f**] final iteration

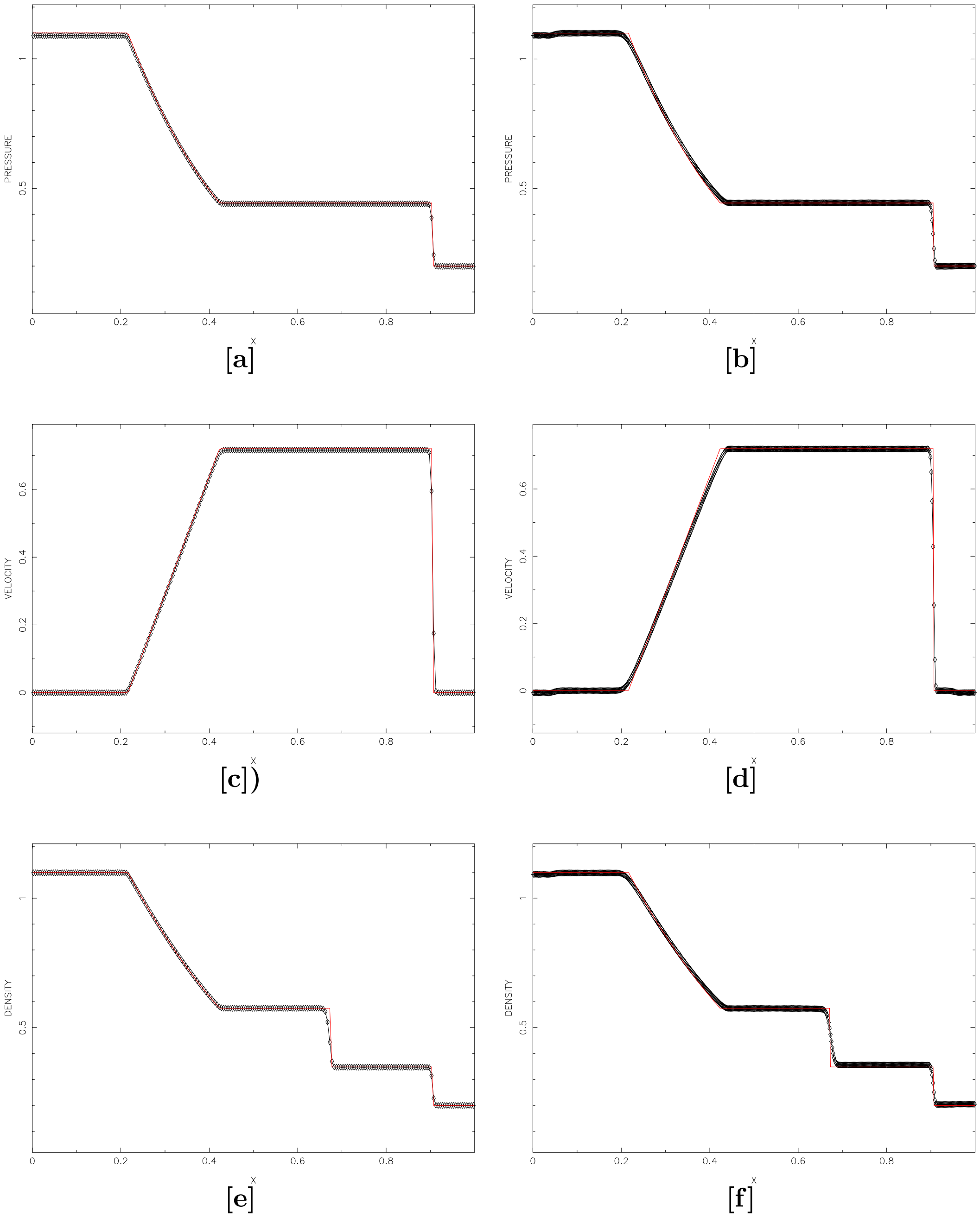


[a] [b]

[c]) [d]

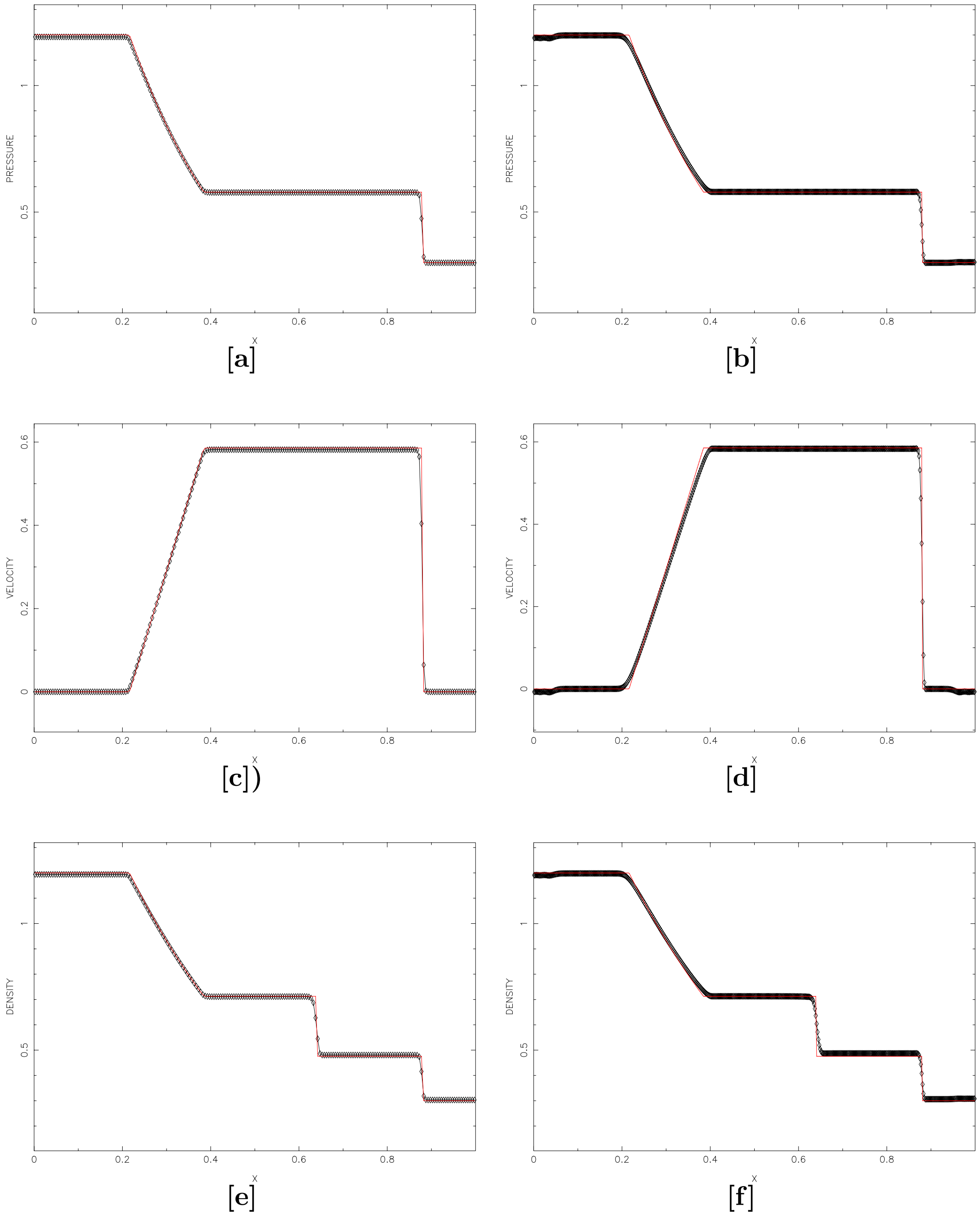
[e] [f]

Figure 10.13. Pressure, density and velocity: numerical ($\diamond$) and analytical (red line) solution for the shock-tube problem at time=0.24 for **HRM** model ([**a**], [**c**] and [**e**]), respectively for the **AVM** model ([**b**], [**d**] and [**f**])



[**a**] [**b**] [**c**]) [**d**] [**e**] [**f**]

Figure 10.14. Pressure, density and velocity: numerical solution (◇) after non smooth optimization **PVAR** and first set of observations (red line) for the shock-tube problem at time=0.24 for **HRM** model ([**a**], [**c**] and [**e**]), respectively for the **AVM** model ([**b**], [**d**] and [**f**])



[**a**] [**b**]

[**c**]) [**d**]

[**e**] [**f**]

Figure 10.15. Pressure, density and velocity: numerical solution ($\diamond$) after non smooth optimization **PVAR** and second set of observations (red line) for the shock-tube problem at time=0.24 for **HRM** model ([**a**], [**c**] and [**e**]), respectively for the **AVM** model ([**b**], [**d**] and [**f**])

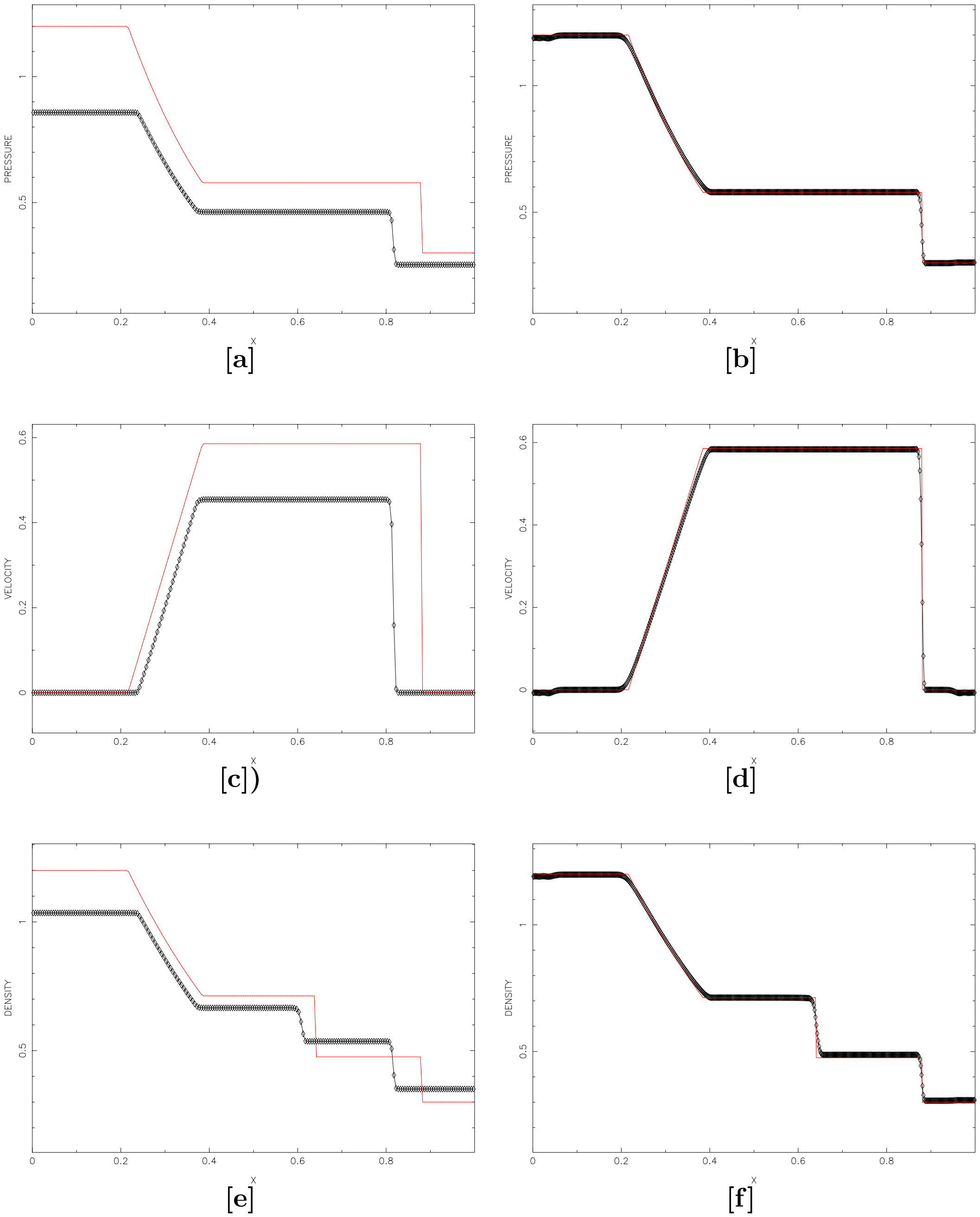
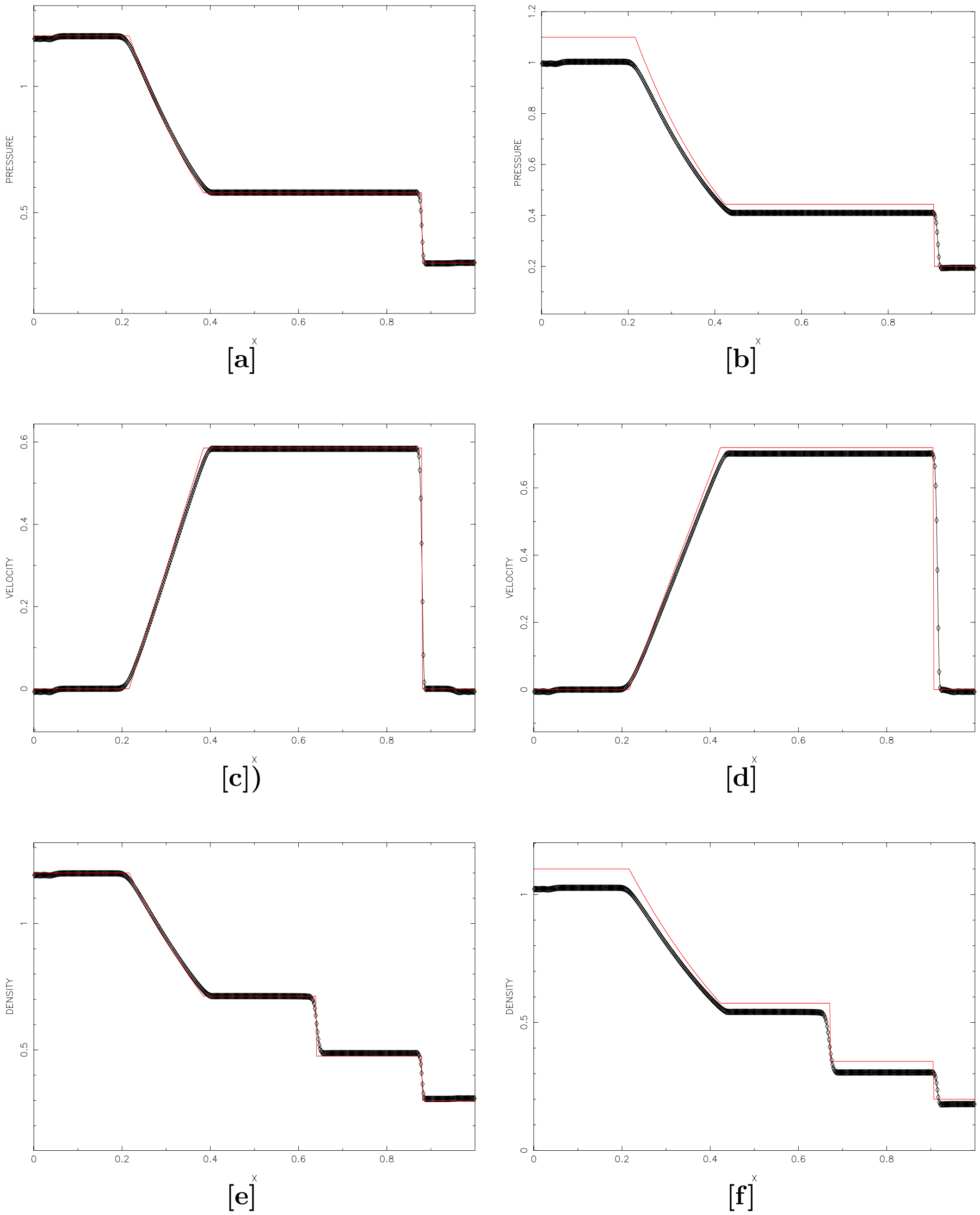


Figure 10.16. Pressure, density and velocity: numerical solution (◇) after **L-BFGS** optimization and second set of observations (red line) for the shock-tube problem at time=0.24 for **HRM** model ([**a**], [**c**] and [**e**]), respectively for the **AVM** model ([**b**], [**d**] and [**f**])

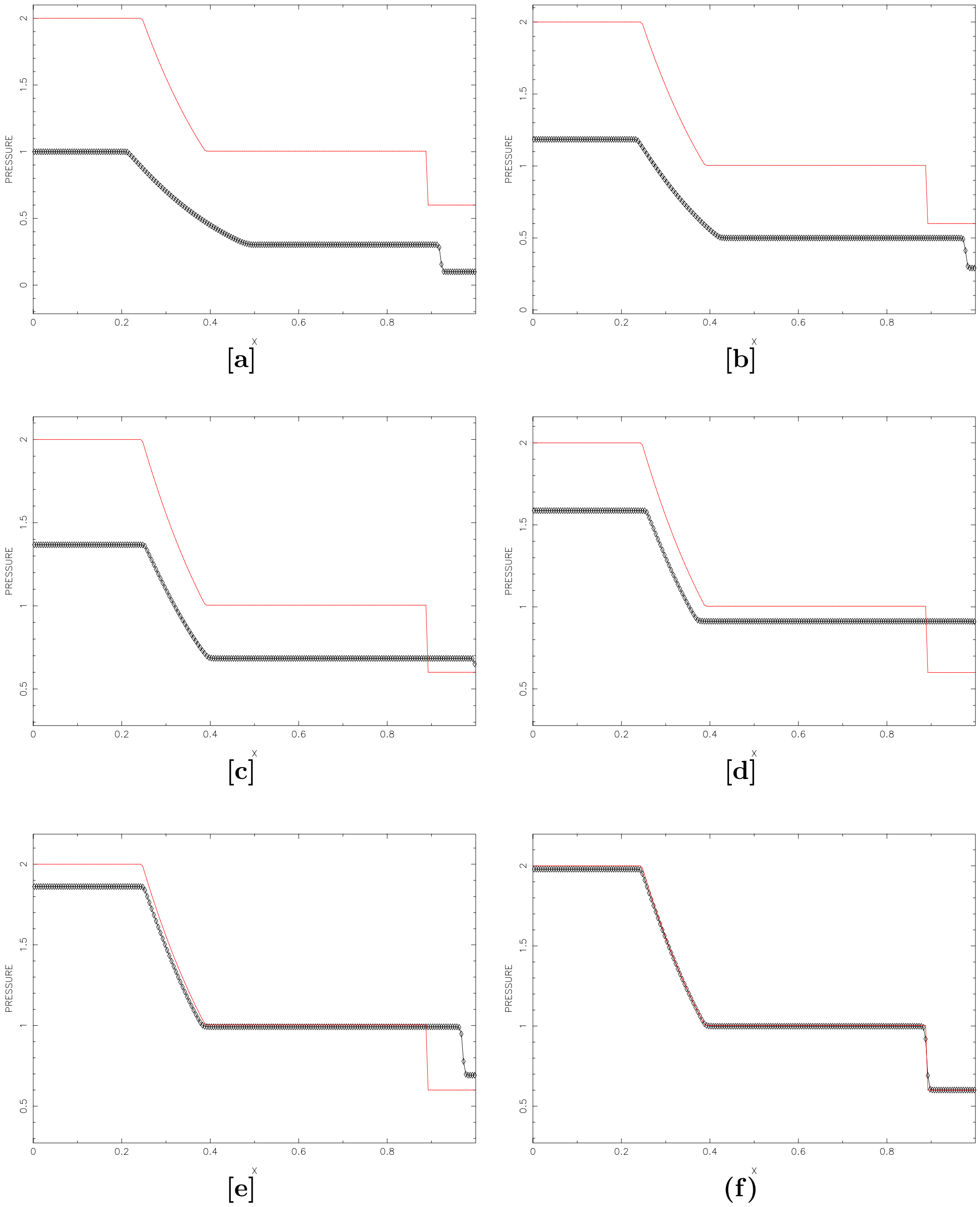


[**a**] [**b**]

[**c**]) [**d**]

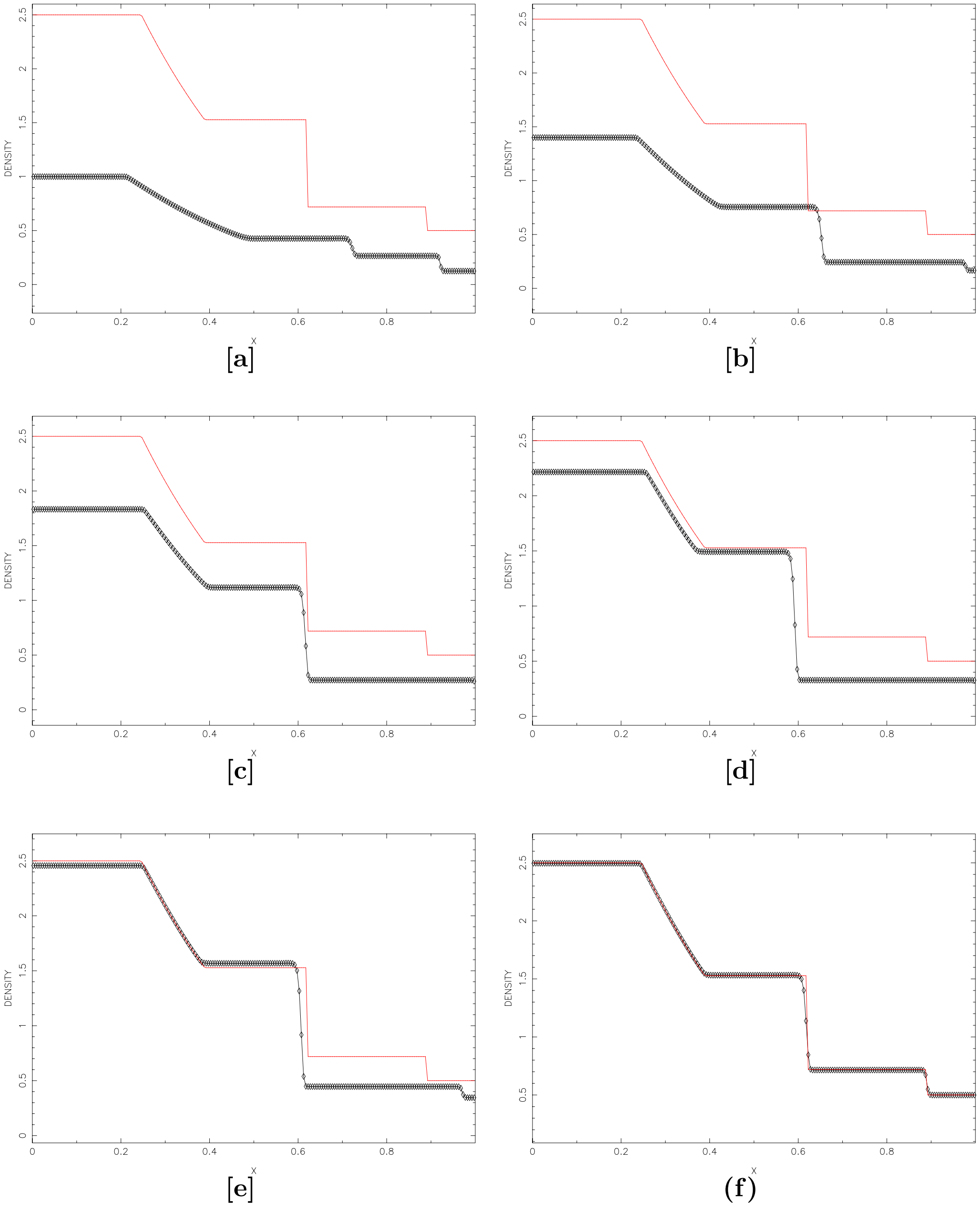
[**e**] [**f**]

Figure 10.17. Pressure, density and velocity: numerical solution after **L-BFGS** optimization ($\diamond$) and first set of observations (red line) for the shock-tube problem at time=0.24 for the **AVM** model for weight=0.0 ([**a**], [**c**] and [**e**]), respectively for no weight considered ([**b**], [**d**] and [**f**])



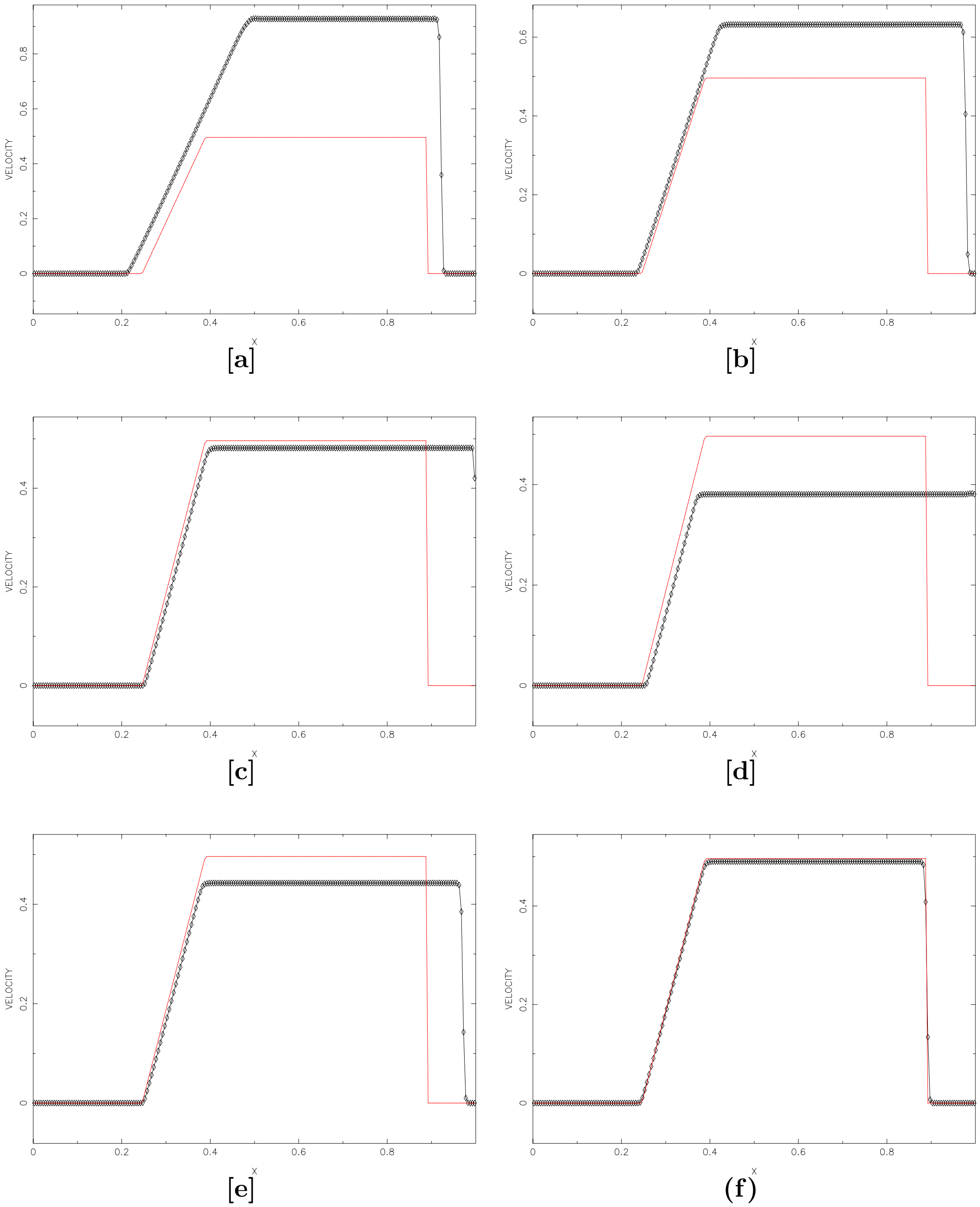
[a] [b] [c] [d] [e] (f)

Figure 10.18. Pressure: numerical solution (⋄) and exact observation (red line) at time=0.24 for the third set of observations and for the **HRM** model during **PVAR** minimization: [**a**] iteration=0; [**b**] iteration=50; [**c**] iteration=100; [**d**] iteration=150; [**e**] iteration=200; [**f**] iteration=268



[a] [b] [c] [d] [e] (f)

Figure 10.19. Density: numerical solution (◇) and exact observation (red line) at time=0.24 for the third set of observations and for the **HRM** model during **PVAR** minimization: [**a**] iteration=0; [**b**] iteration=50; [**c**] iteration=100; [**d**] iteration=150; [**e**] iteration=200; [**f**] iteration=268



[**a**] [**b**]

[**c**] [**d**]

[**e**] (**f**)

Figure 10.20. Velocity: numerical solution (◇) and exact observation (red line) at time=0.24 for the third set of observations and for the **HRM** model during **PVAR** minimization: [**a**] iteration=0; [**b**] iteration=50; [**c**] iteration=100; [**d**] iteration=150; [**e**] iteration=200; [**f**] iteration=268

CHAPTER 11

CONCLUSIONS

We have applied optimal control methods to solve fluid dynamics problems using the adjoint approach for the numerical computation of the gradient (or subgradient) of the cost functional. We investigated differentiable and non differentiable cost functionals which were minimized using smooth or non smooth optimization algorithms.

An optimal control problem of a viscous flow past a roating circular cylinder was chosen for the case of a differentiable cost functional. The 1-D Riemann problem for the Euler equations (shock-tube problem) was considered to exemplify optimal control with a non smooth cost functional. Sensitivity analysis for discontinuous flow was also studied.

Suppression of Karman vortex shedding was achieved for a flow around a rotating cylinder using optimal control. The numerical results obtained agree to a large extent with results obtained by other researchers using other numerical or experimental methods to solve this problem.

An additional result obtained was the significant reduction of the amplitude of the drag coefficient for the flow corresponding to the rotation parameters obtained by the optimal control approach.

The main advantage of the optimal-control approach to flow control is the considerable freedom in choosing the objective function and the parameters of interest. However this approach is very complex and quite demanding computationally.

The adjoint method for computing the gradient of the cost functional with respect to the control parameters provides us with the necessary tool to apply optimal control to the problem of a flow around a rotating cylinder.

Our results were obtained for Reynolds numbers in the range $[60, 1000]$. Future research will apply this method for higher Reynolds numbers, for which there are other regimes, with different characteristics.

Adaptive grid refinement should be considered for improving the accuracy of the results. Another issue, dealing with improving the efficiency of this approach, is to consider the design model version where both forward and adjoint models use parallel programming.

This optimization problem was characterized by its ill-posedness. Our approach for circumventing it was the inclusion of a regularization term in the objective functional. An empirical law for finding suitable penalty parameters was found, allowing efficient minimization to be performed. There are other approaches for dealing with ill-posedness which can be used as well: the utilization of a second-order Tikhonov regularization function (e.g., Alekseev and Navon [4]) or the method of **SVD** (Singular Value Decomposition) which will decompose the problem into well-posed and ill-posed components (e.g., Liu et al. [135], Alekseev and Navon [4]).

Sensitivities are derivatives of the variables or cost functionals that describe the model with respect to parameters that determine the behavior of the model (e.g., initial conditions, boundary conditions or model parameters). They provide information about which of these parameters most influence the model output. We studied sensitivity analysis for a fluid dynamics problem characterized by several types of discontinuities.

Our research was focused on the numerical computation of flow sensitivities with respect to an initial flow parameter for the shock-tube problem (1-D Riemann problem described by Euler equations) for which the exact values of the flow sensitivities are known.

The forward model was chosen such that numerical errors from solving the discontinuities are minimized by a large extent. This was achieved by using an adaptive mesh refinement coupled with a Riemann solver as the discrete forward model. Since the numerical sensitivities are obtained using the tangent linear model (which is derived from the forward model), this implies that we eliminated a majoor source of errors from the numerical values of sensitivities.

Our experience with tangent linear models in higher dimensions (Homescu et al. [101] suggests extending the numerical methodology presented here to higher spatial dimensions. For problems with discontinuities in 2-D or 3-D we expect a decrease in the numerical accuracy of sensitivity computation, compared to the 1-D case. A possible remedy for alleviating this problem was presented by Dadone et al. [44]) and it consists in the application of a smoother to the sensitivities.

Theoretical aspects of linearization for Euler equations were presented. The solution of the linearized system of equations and the sensitivity with respect to a model parameter are solutions of the tangent linear system. The tangent linear model provides a numerical value of the sensitivity which is in better agreement with the analytical solution than any previously published numerical results, to a high extent due to the use of highly accurate adaptive mesh refinement code.

The example chosen for an optimal control problem of a flow with discontinuities was one of flow matching for a 1-D Riemann problem for Euler equations, namely the shock-tube problem, which includes several types of discontinuities. The control variables considered were the initial conditions at the left and at the right of the membrane for pressure and density. Existence results were proved for the solution of the optimal control problem considered here. The cost functional was taken to be the (weighted) difference between the numerical and the *desired* solution of the model. The observations were taken either at the end of the time window or they were time distributed within the assimilation time horizon.

For the present problem flow matching was equivalent to relocation of discontinuities to a desired location. Since in all practical control applications discontinuities are captured using either high-resolution models or models which *smooth* the solution we employed here two numerical models representative of both approaches. For each forward model its corresponding discrete adjoint model was then employed for computing the gradient (or a subgradient) of the cost functional required for carrying out the minimization of the cost functional with respect to the control variables (using either non smooth or smooth algorithms for minimization). The two assimilation windows for minimization were chosen such that the flow with discontinuities retained all its characteristics at the end of each time window. If we were to use a slightly larger time window the model time evolution would change some of the characteristics of discontinuities.

The method of non smooth optimization (**PVAR**) employed for minimizing the cost functional was found to be very robust for our test cases. For each of the different sets of observations employed we obtained optimized values of the control parameters in very good agreement with the *desired* results. The smooth minimization algorithm (**L-BFGS**) provided good results for the shorter time window but failed for the longer time window, even when a scaling the gradient of the cost functional was performed. Better results for **L-BFGS** minimization were obtained if weights were assigned to the points were the shock occurs (the shock points were identified using a method of discontinuities detection). The evolution of the entropy during various stages of the minimization process shows that the numerical solution of the optimal control problem obtained does indeed satisfy the entropy condition. This fact supports our conclusion that the numerical optimal solution is a physical solution, since it is known that the correct weak solution of the shock tube problem must satisfy the entropy condition.

A very useful characteristic of the methodology for optimal control for discontinuous flow presented in this article is the ease with which it can be implemented in applications where the forward model is already discretized (the *discretize-then-differentiate* approach).

Extending this approach to optimal control problems with discontinuities in 2-D or 3-D would render the adjoint method even more appealing computationally, due to the larger number of control parameters involved. It would also apply to more realistic test cases, in particular in aerodynamics (e.g., Jameson [114]).

If the observations are "noisy", one may expect that the cost functional should have new components which will account for the effect of the noise. Both noisy observations and model errors are issues to be addressed in future research.

I plan also to look further into the issue of controllability and observability for the optimal control problem of discontinuous flow. We recall that by *controllability* of the system one means the possibility of influencing independently each state of the system through the inputs; by *observability* of the system one means the possibility of reconstructing each state of the system from the outputs.

An important question which should be addressed in subsequent research is related to the the bounds for the controls. In other words we will try to determine the *desired* values of the flow parameters such that either the optimal control problem cannot be solved theoretically or its numerical solution cannot be found.

We consider our research to be only a small step towards the complete solution of optimal control of problems with discontinuities. Although published results are rather few, one may foresee a growing number of research efforts dedicated to the numerical and theoretical studies of this class of important optimal control problems.

REFERENCES

[1] F. Abergel and R. Temam. On some control problems in fluid mechanics. *Theoretical and Computational Problems in Fluid Dynamics*, 1:303–325, 1990.

[2] K. Afanasiev and M. Hinze. Adaptive control of a wake flow using proper orthogonal decomposition. In J. Cagnol, M.P. Polis, and J.-P. Zolesio, editors, *Shape Optimization and Optimal Design*, pages 317–332. Marcel Dekker, 2001.

[3] L. D. Akulenko. *Problems and Methods of Optimal Control*, volume 286 of *Mathematics and Its Applications*. Kluwer Academic Publishers, 1994.

[4] A. Alekseev and I.M Navon. The analysis of an ill-posed problem using multiscale resolution and second order adjoint techniques. *Computer Methods in Applied Mechanics and Engineering*, 190(15-17):1937–1953, 2001.

[5] J.D. Anderson. *Modern compressible flow, with historical perspective*. Mc-Graw Hill in mechanical engineering. Mc-Graw Hill, 1990.

[6] W.K. Anderson and A. Venkatakrishnan. Aerodynamic design optimization on unstructured grids with a continuous adjoint formulation. *Computers and Fluids*, 28(4-5):443–480, 1999.

[7] E. Arian and M.D. Salas. Admitting the inadmissible: adjoint formulation for incomplete cost functionals in aerodynamic optimization. *AIAA Journal*, 37(1):37–45, 1999.

[8] H. Badr, M. Coutanceau, S. Dennis, and C. Menard. Sur la comparaison des calculs numeriques et des visualisations de l'ecoulement d'un fluide visqueux engendre par un cylindre en translation et rotation. *Comptes Rendus de L'Academie des Sciences Paris*, 300(12):529–533, 1985.

[9] H. Badr, M. Coutanceau, S. Dennis, and C. Menard. Unsteady flow past a rotating circular cylinder at Reynolds numbers 1000 and 10000. *Journal of Fluid Mechanics*, 220:459–484, 1990.

[10] S. Baek and H. Sung. Numerical simulation of the flow behind a rotary oscillating circular cylinder. *Physics of Fluids*, 10(4):869–876, 1998.

[11] H.M. Barkla and L.J. Auchterlonie. The Magnus or Robins effect on rotating spheres. *Journal of Fluid Mechanics*, 47(3):437–447, 1971.

[12] P.I. Barton, J.R. Banga, and S. Galan. Optimization of hybrid discrete/continuous dynamic systems. *Computers and Chemical Engineering*, 24(9-10):2171–2182, 2000.

[13] T. Bein, H. Hanselka, and E. Breitbach. An adaptive spoiler to control the transonic shock. *Smart Materials and Structures*, 9(2):141–148, 2000.

[14] P. Berggren. Numerical solution of a flow control problem: vorticity reduction by dynamic boundary action. *Siam Journal in Control and Optimization*, 19(3):829–860, 1998.

[15] R. S. Berry, V. Kazakov, S. Sieniutycz, Z. Szwast, and A. M. Tsirlin. *Thermodynamic Optimization of Finite-Time Processes*. Wiley, 2000.

[16] M. Berz, C. Bischof, G. Corliss, and A. Griewank (Eds.). *Computational differentiation: Techniques, Applications and Tools*, volume 89 of *Proceedings in Applied Mathematics*. SIAM, 1996.

[17] T. Bewley, R. Temam, and M. Ziane. A general framework for robust control in fluid mechanics. *Physica D*, 138:360–392, 2000.

[18] J. Birkemeyer, H. Rosemann, and E. Stanewsky. Shock control on a swept wing. *Aerospace Science and Technology*, 4(3):147–156, 2000.

[19] L.G. Birta and T.I. Oren. A robust procedure for discontinuity handling in continuous system simulation. *Transactions of the Society for Computer Simulation*, 2(3):189–205, 1985.

[20] J.F. Bonnans, J. Gilbert, C. Lemarechal, and C. Sagastizabal. *Optimisation numerique: aspects theoriques et pratiques*, volume 27 of *Mathematiques et Applications*. Springer-Verlag, Paris, 1997.

[21] F. Bouchut and F. James. Differentiability with respect to initial data for a scalar conservation law. In M. Fey and R. Jeltsch, editors, *Proceedings of the Seventh International Conference on Hyperbolic Problems: Theory, Numerics, Applications (ETH Zurich, February 9-13, 1998)*, number 129 in International Series of Numerical Mathematics, pages 113–118. Birkhauser, 1999.

[22] G.W. Burgreen and O. Baysal. Three dimensional aerodynamic shape optimization using discrete sensitivity analysis. *AIAA Journal*, 34(9):1761–1770, 1996.

[23] J.A. Burns and Y.R. Ou. Effect of rotation rate on the forces of a rotating cylinder: simulation and control. ICASE Report 93-11, ICASE, 1993.

[24] D.G. Cacuci. Sensitivity theory for nonlinear systems: I. Nonlinear functional analysis approach. *Journal of Mathematical Physics*, 22(12):2794–2802, 1981.

[25] D.G. Cacuci. Sensitivity theory for nonlinear systems: II. Extensions to additional classes of responses. *Journal of Mathematical Physics*, 22(12):2803–2812, 1981.

[26] D.G. Cacuci. Global optimization and sensitivity analysis. *Nuclear Science and Engineering*, 104(1):78–88, 1990.

[27] I. Charpentier. Checkpointing schemes for adjoint codes: application to the meteorological model Meso-NH. *SIAM Journal on Scientific Computing*, 22(6):2135–2151, 2001.

[28] Y. Chen, Y.R. Ou, and A. Pearlstein. Development of the wake behind a circular cylinder impulsively started into rotatory and rectilinear motion. *Journal of Fluid Mechanics*, 253:449–484, 1993.

[29] Y.T. Chew, M. Cheng, and S.C. Luo. A numerical study of flow past a rotating circular cylinder using a hybrid vortex scheme. *Journal of Fluid Mechanics*, 299:35–71, 1995.

[30] H. Choi, M. Hinze, and K. Kunisch. Instantaneous control of backward-facing step flows. *Applied Numerical Mathematics*, 31(2):133–158, 1999.

[31] M.H. Chou. Synchronization of vortex shedding from a cylinder under rotary oscillation. *Computers and Fluids*, 26(8):755–774, 1997.

[32] M.H. Chou. Numerical study of vortex shedding from a rotating cylinder immersed in a uniform flow field. *International Journal for Numerical Methods in Fluids*, 32(5):545–567, 2000.

[33] G.S. Christensen, M.E. El-Hawary, and S.A. Soliman. *Optimal control applications in electric power systems*, volume 35 of *Mathematical concepts and methods in science and engineering*. Plenum Press, 1987.

[34] G.S. Christensen, S.A. Solimannd, and R. Nieva. *Optimal control of distributed nuclear reactors*, volume 41 of *Mathematical concepts and methods in science and engineering*. Plenum Press, 1990.

[35] E.M. Cliff, M. Heinkenschloss, and A.R. Shenoy. On the optimality system for a 1-D Euler flow problem. AIAA 96-3993, 1996.

[36] E.M. Cliff, M. Heinkenschloss, and A.R. Shenoy. An optimal control problem for flows with discontinuities. *Journal of Optimization Theory and Applications*, 94(2):273–309, 1997.

[37] E.M. Cliff, M. Heinkenschloss, and A.R. Shenoy. Adjoint-based methods in aerodynamic design optimization. In J. Borggaard, J. Burns, E. Cliff, and S. Schreck, editors, *Computational methods for optimal design and control Proceedings of the AFSOR Workshop on Optimal Design and Control, Arlington, VA, 30 September – 3 October 1997*, pages 91–112. Birkhauser, 1998.

[38] E.M. Cliff, M. Heinkenschloss, and A.R. Shenoy. Airfoil design by an All-At-Once method. *International Journal for Computational Fluid Mechanics*, 11:3–25, 1998.

[39] J. Coron. On the controllability of the 2-D incompressible Navier-Stokes equations with the Navier slip boundary equations. *ESAIM: Control, optimization and calculus of variations*, 1:35–75, 1996.

[40] D. Cossin and F. M. Aparicio. *Optimal control of credit risk*, volume 3 of *Advances in computational management science*. Kluwer Academic Publishers, 2001.

[41] T.J. Cowan. Private communication, 2001.

[42] A. Dadone and B. Grossman. Fast convergence of inviscid fluid dynamic design problems. Proceedings of ECCOMAS 2000 Barcelona 11-14 September 2000 `http://www.imamod.ru/jour/conf/ECCOMAS2000/pdf/148.pdf`, 2000.

[43] A. Dadone and B. Grossman. Progressive optimization of inverse fluid dynamic design problems. *Computers and Fluids*, 29(1):1–32, 2000.

[44] A. Dadone, M. Valorani, and B. Grossman. Optimization of 2D fluid design problems with nonsmooth or noisy objective function. In J.A. Desideri, J. Hirsch, E. Stein, J. Periaux, M. Pandolfi, P. Le Tallec, and E. O'Nate, editors, *Computational Fluid Dynamics '96 Proceedings of ECCOMAS 96, Paris, France, September 9-13, 1996*, pages 425–430. Wiley and Sons, 1996.

[45] J.E. Dennis, M. Heinkenschloss, and L.N. Vicente. TRICE: trust-Region Interior-Point algorithms for optimal control and engineering design problems. `http://www.caam.rice.edu/~trice`.

[46] S. Dennis, P. Nguyen, and S. Rocabiyik. The flow induced by a rotationally oscillating and translating circular cylinder. *Journal of Fluid Mechanics*, 407:123–144, 2000.

[47] N. DiCesare and O. Pironneau. Shock sensitivity analysis. *Computational Fluid Dynamics Journal*, 9(2):1–6, 2000.

[48] Y. Ding and M. Kawahara. Secondary instabilities of wakes of a circular cylinder using a finite element method. *International Journal of Computational Fluid Dynamics*, 13(3):279–312, 2000.

[49] F. Dubois and G. Mehlman. A non-parametrized entropy fix for Roe's method. *AIAA Journal*, 31(1):199–200, 1993.

[50] T. F. Edgar, D. M. Himmelblau, and L. S. Lasdon. *Optimization of chemical processes*. McGraw-Hill chemical engineering series. McGraw-Hill, 2001.

[51] E.Godlewski and P.A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*, volume 118 of *Applied Mthematical Sciences*. Springer-Verlag, 1996.

[52] L. Elden. Algorithms for the regularization of ill-conditioned least squares problems. *BIT*, 17:134–145, 1977.

[53] G. Erlebacher and M.Y. Hussaini. Shock-shape alteration caused by interaction with organized structures. *AIAA Journal*, 38(6):1002–1009, 2000.

[54] G. Erlebacher, M.Y. Hussaini, and T.L. Jackson. Nonlinear strong shock interactions: a shock-fitted approach. *Theoretical and Computational Fluid Dynamics*, 11(1):1–29, 1998.

[55] G. Erlebacher, M.Y. Hussaini, and C.W. Shu. Interaction of a shock with a longitudinal vortex. *Journal of Fluid Mechanics*, 337:129–153, 1997.

[56] H.E. Fleming. Equivalence of regularization and truncated iteration in the solution of ill-posed image reconstruction problem. *Linear Algebra and its Applications*, 130:133–150, 1990.

[57] R. Fletcher. *Practical methods of optimization.* Wiley and Sons, 1987.

[58] P.D. Frank and G.Y. Shubin. A comparison of optimization-based approaches for a model computational aerodynamics design problem. *Journal of Computational Physics*, 98(1):74–89, 1992.

[59] A. Fursikov, M. Gunzburger, and L. Hou. Boundary value problems and optimal boundary control for the Navier-Stokes system: the two-dimensional case. *SIAM Journal of Control and Optimization*, 36(3):852–894, 1998.

[60] M. Gad-El-Hak. Modern developments in flow control. *Applied Mechanics Reviews*, 49:365–379, 1996.

[61] M. Gad-El-Hak. *Flow control: Passive, active and reactive flow management.* Cambridge University Press, 2000.

[62] T.B. Gatski, M.Y. Hussaini, and J.L. Lumley. *Simulation and modeling of turbulent flows.* ICASE/LaRC Series in Computational science and Engineering. Oxford University Press, 1996.

[63] O. Ghattas and J. Bark. Optimal control of 2-D and 3-D incompressible Navier-Stokes flows. *Journal of Computational Physics*, 136(2):231–244, 1997.

[64] M.B. Giles and N.A. Pierce. Adjoint equations in CFD: duality, boundary conditions and solution behaviour. AIAA 97-1850, 1997.

[65] M.B. Giles and N.A. Pierce. On the properties of solutions of the adjoint euler equations. In M. Baines, editor, *Numerical Methods for Fluid Dynamics VI.* ICFD, Oxford, 1998.

[66] M.B. Giles and N.A. Pierce. An introduction to the adjoint approach to design. *Flow Turbulence and Combustion*, 65(3-4):393–415, 2000.

[67] M.B. Giles and N.A. Pierce. Analytic adjoint solutions for the quasi-1D Euler equations. *Journal of Fluid Mechanics*, 426:327–345, 2001.

[68] E. Gillies. Low dimensional control of the circular cylinder wake. *Journal of Fluid Mechanics*, 371:157–178, 1998.

[69] A. A. Giunta and J. Sobieszczanski-Sobieski. Progress toward using sensitivity derivatives in a high fidelity aeroelastic analysis of a supersonic transport. AIAA 98-4763, 1998.

[70] E. Giusti. *Minimal surfaces and functions of bounded variation.* Birkhauser Verlag, 1984.

[71] E. Godlewski and P.A. Raviart. The linearized stability of solutions of nonlinear hyperbolic systems of conservation laws. A general numerical approach. *Mathematics and Computers in Simulation*, 98:77–95, 1999.

[72] C.H. Goldsmith. Sensitivity analysis. In P. Armitage and T. Colton, editors, *Encyclopedia of Biostatistics*. Wiley, 1998.

[73] G.H. Golub, M.T. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.

[74] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, 1996.

[75] C. Gourieroux and J. Jasiak. Nonlinear innovations and impulse responses. `http://dept.econ.yorku.ca/jasj/papers.html`, 2000.

[76] J.M.R. Graham. Report on the session comparing computation of flow past circular cylinders with experimental data. In H. Eckelmann, J.M.R. Graham, P. Huerre, and P.A. Monkewitz, editors, *Procedings of IUTAM Symposium: Bluff-Body wakes, Dynamics and Instabilities, 7-11 September 1992 Gottingen*. Springer Verlag, 1993.

[77] W.R. Graham, J. Peraire, and K.Y. Tang. Optimal control of vortex-shedding using low order models. Part I: Open-loop model development. *International Journal of Numerical Methods in Engineering*, 44(7):950–972, 1999.

[78] W.R. Graham, J. Peraire, and K.Y. Tang. Optimal control of vortex-shedding using low order models. Part II: Model based control. *International Journal of Numerical Methods in Engineering*, 44(7):973–990, 1999.

[79] M. Griebel, T. Dornseifer, and T. Neunhoeffer. *Numerical simulation in fluid dynamics*, volume 3 of *SIAM Monographs on Mathematical Modeling and Computation*. SIAM, Philadelphia, 1998.

[80] A. Griewank. On automatic differentiation. In M.M. Iri and K. Tanabe, editors, *Mathematical Programming: Recent developments and applications*, pages 83–108. Kluwer, 1989.

[81] A. Griewank. Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation. *Optimization Methods and Software*, 1(1):35–54, 1992.

[82] A. Griewank. *Evaluating derivatives, principles and techniques of algorithmic differentiation*, volume 19 of *Frontiers in Applied Mathematics*. SIAM, 2000.

[83] A. Griewank and G. Corliss (Eds.). *Automatic differentiation of algorithms: Theory, implementation and algorithms*. SIAM, 1991.

[84] A. Griewank and A. Walther. Algorithm 799: Revolve: an implementation of checkpointing for the reverse or adjoint mode of computational derivatives. *ACM Transactions on Mathematical Software*, 26(1):19–45, 2000.

[85] M. Gunzburger. Sensitivities in computational methods for optimal flow control. In J. Borggaard, J. Burns, E. Cliff, and S. Schreck, editors, *Computational methods for optimal design and control*, Progress in Systems and Control Theory, pages 197–236. Springer Verlag, 1996.

[86] M. Gunzburger. Sensitivities, adjoints and flow optimization. *International Journal for Numerical Methods in Fluids*, 31(1):53–78, 1999.

[87] M. Gunzburger, L. Hou, and T. Svobodny. Analysis and finite element approximation of optimal control problems for the stationary Navier-Stokes equations with distributed and Neumann controls. *Mathematics of Computation*, 57(195):123–151, 1991.

[88] M. Gunzburger, L. Hou, and T. Svobodny. Boundary velocity control of incompressible flow with an application to viscous drag reduction. *SIAM Journal of Control and Optimization*, 30(1):167–181, 1992.

[89] M. Gunzburger and H. Lee. Feedback control of Karman vortex shedding. *Transactions of ASME -Journal of Applied Mechanics*, 63:828–835, 1996.

[90] M. Gunzburger and S. Manservisi. The velocity tracking problem for Navier-Stokes flows with bounded distributed controls. *SIAM Journal of Control and Optimization*, 37(6):1913–1945, 1999.

[91] M. Gunzburger(Ed.). *Flow Control*, volume 68 of *The IMA Volumes in Mathematics and Its Applications*. Springer Verlag, 1996.

[92] A. Habbal. Direct approach to the minimization of the maximal stress over an arch structure. *Journal of Optimization Theory and Applications*, 97(3):551–578, 1998.

[93] A. Habbal. Nonsmooth shape optimization applied to linear acoustics. *SIAM Journal on Optimization*, 8(4):989–1006, 1998.

[94] P.C. Hansen. *Rank-deficient and discrete ill-posed problems. Numerical aspects of linear inversion*, volume 4 of *SIAM Monographs on Mathematical Modelling and Computation*. SIAM, Philadelphia, 1998.

[95] P.C. Hansen and D.P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.

[96] A. Harten and J.M. Hyman. Self adjusting grid methods for one-dimensional hyperbolic conservation laws. *Journal of Computational Physics*, 50(2):235–269, 1983.

[97] E. Hassold. Automatic differentiation applied to a nonsmooth optimization problem. In J.-A. Desideri, P. Le Tallec, E. Onate, J. Periaux, and E. Stein, editors, *Numerical Methods in Engineering '96*, pages 835–841. Wiley, 1996.

[98] J.W. He, R. Glowinski, R. Metcalfe, A. Nordlander, and J. Periaux. Active control and drag optimization for flow past a circular cylinder I: Oscillatory cylinder. *Journal of Computational Physics*, 163(1):83–117, 2000.

[99] J.C. Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety*, 42(2-3):327–367, 1993.

[100] C. Homescu and I.M. Navon. Numerical and theoretical considerations for sensitivity calculation of discontinuous flow. *Accepted for publication in Systems and Control Letters*, 2002.

[101] C. Homescu, I.M. Navon, and Z. Li. Suppression of vortex shedding for flow around a circular cylinder using optimal control. *International Journal for Numerical Methods in Fluids*, 38(1):43–69, 2002.

[102] L. Hou and S.S. Ravindran. A penalized Neumann control approach for solving an optimal Dirichlet control problem for the Navier-Stokes equations. *Siam Journal of Control and Optimization*, 36(5):1795–1814, 1998.

[103] L. Hou and S.S. Ravindran. Penalty methods for numerical approximations of optimal boundary flow control problems. *International Journal of Computational Fluid Dynamics*, 11(1-2):157–167, 1998.

[104] L. Hou, S.S. Ravindran, and Y. Yan. Numerical solutions of optimal distributed control problems for incompressible flows. *International Journal of Computational Fluid Dynamics*, 8(2):99–114, 1997.

[105] L. Hou and Y. Yan. Dynamics and approximations of a velocity tracking problem for the Navier-Stokes flows with piecewise distributed controls. *SIAM Journal on Control and Optimization*, 35(6):1847–1885, 1997.

[106] X. Huang. Feedback control of vortex shedding from a circular cylinder. *Experiments in Fluids*, 20(3):218–224, 1996.

[107] M.Y. Hussaini and G. Erlebacher. Interaction with an entropy spot with a shock. *AIAA Journal*, 37(3):346–356, 1999.

[108] L. Huyse and M.R. Lewis. Aerodynamic shape optimization of two-dimensional airfoils under uncertain conditions. ICASE Report 2001-1, ICASE, 2001.

[109] A. Iollo and M.D. Salas. Contribution to the optimal shape design of two-dimensional internal flows with embedded shocks. *Journal of Computational Physics*, 125(1):124–134, 1996.

[110] A. Iollo and M.D. Salas. Optimum transonic airfoils based on Euler equations. *Computers and Fluids*, 28(4-5):653–674, 1999.

[111] A. Iollo, M.D. Salas, and S. Ta'asan. Shape optimization governed by the Euler equations using an adjoint method. ICASE Report 93-78, ICASE, 1993.

[112] K. Ito and S.S. Ravindran. A reduced-order method for simulation and control of fluid flows. *Journal of Computational Physics*, 143(2):403–425, 1998.

[113] F. James and M. Sepulveda. Convergence results for the flux identification in a scalar conservation law. *SIAM Journal on Control and Optimization*, 37(3):869–891, 1999.

[114] A. Jameson. A perspective on computational algorithms for aerodynamic analysis and design. *Progress in Aerospace Sciences*, 37(2):197–243, 2001.

[115] P.H.M. Janssen, W.A. Rossing, and J. Rotmans. Uncertainty analysis and sensitivity analysis: an inventory of ideas, methods and techniques from the literature. RIWM Report 958805001, RIVM, Bilthoven, The Netherlands, 1990.

[116] R. Joslin, M. Gunzburger, R. Nicolaides, G. Erlebacher, and M.Y. Hussaini. Self-contained automated methodology for optimal flow control. *AIAA Journal*, 35(5):816–824, 1997.

[117] H. Juarez, R. Scott, R. Metcalfe, and B. Bagheri. Direct simulation of freely rotating cylinders in viscous flows by high-order finite element methods. *Computers and Fluids*, 29(5):547–582, 2000.

[118] S. Kang, H. Choi, and S. Lee. Laminar flow past a rotating circular cylinder. *Physics of Fluids*, 11(11):3312–3321, 1999.

[119] T. von Karman. On the mechanism of drag generation on the body moving in fluid . Part I and II (in German). *Nachrichten Gesellschaft Wissenschaften, Gottingen*, pages 509–517 and 547–556, 1911.

[120] M.E. Kilmer and D.P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM Journal on Matrix Analysis and Applications*, 22(4):1204–1221, 2001.

[121] K.C. Kiwiel. *Methods of descent for nondifferentiable optimization*, volume 1133 of *Lecture Notes in Mathematics*. Springer, 1985.

[122] K.C. Kiwiel. Proximity control in bundle methods for convex nondifferentiable minimization. *Mathematical Programming*, 46:105–122, 1990.

[123] K. Kwon and H. Choi. Control of laminar vortex shedding behind a circular cylinder using splitter plates. *Physics of Fluids*, 10:479–485, 1996.

[124] D. Lee and T. Pavlidis. One-dimensional regularization with discontinuities. *IEEE transactions on pattern analysis and machine intelligence*, 10(6):822–828, 1988.

[125] C. Lemarechal. Nondifferentiable optimization, Subgradient and ϵ subgradient methods. In *Optimization and Operations Research*, number 117 in Lecture Notes in Economics and Mathematical Systems, pages 191–199. Springer-Verlag, 1976.

[126] C. Lemarechal. Nondifferentiable optimization. In G.L. Nemhauser, A.H.G. Rinnooy Kan, and M.J. Todd, editors, *Handbooks in Operations Research and Management Science. Volume 1: Optimization*, pages 529–572. North Holland, 1989.

[127] D. Leonard and N.V. Long. *Optimal control theory and static optimization in economics*. Cambridge University Press, 1992.

[128] R. Leveque. CLAWPACK: A software package for conservation laws and hyperbolic systems. `http://www.amath.washington.edu/~rjl/clawpack`.

[129] R. Leveque. *Numerical methods for conservation laws.* Lectures in Mathematics (ETH Zurich). Birkhauser, 1992.

[130] S. Li. *Adaptive mesh methods and software for time-dependent partial differential equations.* PhD thesis, Dept. Of Computer Science, University of Minnesota, 1998.

[131] Z. Li, I.M. Navon, M.Y. Hussaini, and F.X. LeDimet. Optimal control of cylinder wakes via suction and blowing. *Computers and Fluids*, 31(8):In press (available online), 2002.

[132] H.W. Liepmann and A. Roshko. *Elements of gas dynamics*, volume 8 of *Galcit aeronautical series.* Wiley and Sons, 1957.

[133] G.P Ling and T.M Shih. Numerical study on the vortex motion patterns around a rotating circular cylinder and their critical characters. *International Journal for Numerical Methods in Fluids*, 29(2):229–248, 1999.

[134] D. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming B*, 45:503–528, 1989.

[135] J. Liu, B. Guerrier, and C. Benard. A sensitivity decomposition for the regularized solution of the inverse heat conduction problems by wavelets. *Inverse Problems*, 11(6):1177–1187, 1995.

[136] L. Luksan and J. Vlcek. NDA: Algorithms for nondifferentiable optimization. ICS Report 797, Institute of Computer Science, Academy of Sciences of the Czech Republic, 2000.

[137] L. Luksan and J. Vlcek. Algorithm811: NDA: Algorithms for nondifferentiable optimization. *ACM Transactions on Mathematical Software*, 27(2):193–214, 2001.

[138] M.M. Makela and P. Neittaanmaki. *Nonsmooth Optimization.* World Scientific, 1992.

[139] G. Mao and L.R. Petzold. Efficient integration over discontinuities for differential-algebraic systems. *Computers and Mathematics with Applications*, 43(1-2):65–79, 2002.

[140] T. Matsuzawa and M. Hafez. Optimum shape design using adjoint equations for compressible flows with shock waves. *International Journal for Computational Fluid Dynamics*, 7(3):343–365, 1998.

[141] T. Matsuzawa and M. Hafez. Treatment of shock waves in design optimization via adjoint equation approach. *International Journal for Computational Fluid Dynamics*, 7(4):405–425, 1998.

[142] L.W. Mays. *Optimal control of hydrosystems.* Marcel Dekker, 1997.

[143] G. Metivier. Stability of multidimensional shocks. In T.P. Liu, H. Freistuhler, and A. Szepessy, editors, *Advances in the Theory of Shock Waves*, pages 25–103. Birkhauser, 2001.

[144] V.J. Modi. Moving surface boundary-layer control: A review. *Journal of Fluids and Structures*, 11(6):627–663, 1997.

[145] B. Mohammadi and O. Pironneau. *Applied Shape Optimization for Fluids.* Numerical Mathematics and Scientific Computation. Oxford Science Publications, 2001.

[146] V.M. Morozov. On the solution of functional equations by the method of regularization. *Soviet Mathematics Doklady*, 7:414–417, 1966.

[147] R.P. Narducci, B. Grossman, and R.T Haftka. Sensitivity algorithms for an inverse design problem involving a shock wave. *Inverse Problems in Engineering*, 2:49–83, 1995.

[148] I.M. Navon, X. Zou, J. Derber, and J. Sela. Variational data assimilation with an adiabatic version of the NMC spectral model. *Monthly Weather Review*, 120(7):1433–1446, 1992.

[149] J.C. Newman, A.C. Taylor, R.W. Barnwell, P. Newman, and G.J.W. Hou. Overview of sensitivity analysis and shape optimization for complex aerodynamic configurations. *AIAA Journal*, 36(1):87–95, 1999.

[150] T. V. Nguyen, A. Devgan, and O. J. Nastov. Adjoint transient sensitivity computation in piecewise linear simulation. In *Proceedings of the 35th Design Automation Conference, San Francisco, June 1998*, pages 477–482, `http://citeseer.nj.nec.com/nguyen98adjoint.html`, 1998.

[151] J. Nocedal. Updating Quasi-Newton matrices with limited storage. *Mathematics of Computation*, 35(151):773–782, 1980.

[152] J. Nocedal and S.J. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer, 1999.

[153] D.P. O'Leary and J.A. Simmons. A bidiagonalization-regularization procedure for large scale discretization of ill-posed problems. *SIAM Journal on Scientific and Statistical Computing*, 2(4):474–489, 1981.

[154] Y.R. Ou. Mathematical modeling and numerical simulation in external flow control. In M. Gunzburger, editor, *Flow Control*, pages 218–253. Springer Verlag, 1996.

[155] A. Oyama, S. Obayashi, and K. Nakahashi. Transonic wing optimization using genetic algorithms. AIAA 97-1854, 1997.

[156] S. Ozono. Flow control of vortex shedding by a short splitter plate asymmetrically arranged downstream of a cylinder. *Physics of Fluids*, 11(10):2928–2934, 1999.

[157] D. Park, D. Ladd, and E. Hendricks. Feedback control of von Karman vortex shedding behind circular cylinders at low Reynolds numbers. *Physics of Fluids*, 6(7):2390–2405, 1994.

[158] T. Park and P.I. Barton. State event location in differential-algebraic models. *ACM Transactions on Modelling and Computer Simulation*, 6(2):137–165, 1996.

[159] A. Pentek and J. Kadtke. Dynamical control for capturing vortices near bluff bodies. *Physical Review E*, 58(2):1883–1898, 1998.

[160] M. Piasecki and N. Katopodes. Control of contaminant releases in rivers. I: adjoint sensitivity analysis. *Journal of Hydraulic Engineering*, 123(6):488–492, 1997.

[161] L. Prandtl. Bericht ueber untersuchungen zur ausgebileten turbulenz. *ZAMM*, 3:136–139, 1925.

[162] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C. The Art of Scientific Computing 2nd ed.* Cambridge University Press, 1997.

[163] J. Peraire R. Lohner, K. Morgan and M. Vahdati. Finite element flux-corrected transport (FEM-FCT) for the Euler and Navier-Stokes equations. *International Journal for Numerical Methods in Fluids*, 7:1093–1109, 1987.

[164] J.M. Restrepo, G.K. Leaf, and A. Griewank. Circumventing storage limitations in variational data assimilation studies. *SIAM Journal on Scientific Computing*, 19(5):1586–1605, 1998.

[165] P.L. Roe. Approximate Riemann solvers, parameter vectors and difference schemes. *Journal of Computational Physics*, 43:357–372, 1981.

[166] P.L. Roe. Sonic flux formulae. *SIAM Journal on scientific and statistical computing*, 13(2):611–630, 1992.

[167] H. Sakamoto and H. Haniu. Optimum suppression of fluid forces acting on a circular cylinder. *Journal of Fluids Engineering-Transactions of the ASME*, 116(2):221–227, 1994.

[168] B.F. Sanders and N.D. Katopodes. Control of canal flow by adjoint sensitivity method. *Journal of Irrigation and Drainage Engineering*, 125(5):287–297, 1999.

[169] B.F. Sanders and N.D. Katopodes. Adjoint sensitivity analysis for shallow water wave control. *Journal of Engineering Mechanics*, 126(9):909–919, 2000.

[170] A. Satelli, K. Chan, and E.M. Scott (Eds.). *Sensitivity Analysis.* Wiley Series in Probability and Statistics. Wiley, 2000.

[171] H. Schramm and J. Zowe. A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM Journal on Optimization*, 2(1):121–152, 1992.

[172] Atle Seierstad and Knut Sydsaeter. *Optimal control theory with economic applications*, volume 24 of *Advanced textbooks in economics.* North-Holland, 1987.

[173] S. P. Sethi and G. L. Thompson. *Optimal control theory: applications to management science and economics.* Kluwer Academic Publishers, 2000.

[174] F. Shakib, T.J.R. Hughes, and Z. Johan. A new finite formulation for computational fluid dynamics: X. the compressible Euler and Navier-Stokes equation. *Computer Methods in Applied Mechanics and Engineering*, 89:141–219, 1991.

[175] N.Z. Shor. *Minimization methods for Non-differentiable functions*, volume 3 of *Springer Series in Computational Mathematics*. Springer, 1985.

[176] George M. Siouris. *An Engineering Approach to Optimal Control and Estimation Theory*. Wiley, 1996.

[177] G.A. Sod. A survey of finite-difference methods for systems of nonlinear conservation laws. *Journal of Computational Physics*, 27(1):1–31, 1978.

[178] S.S. Sritharan. An optimal control problem in exterior hydrodynamics. *Proceedings of the Royal Society of Edinburgh*, 121A:5–32, 1992.

[179] S.S. Sritharan(Ed.). *Optimal control of viscous flow*. SIAM Press, 1998.

[180] E. Stanewsky. Adaptive wing and flow control technology. *Progress in Aerospace Sciences*, 37(7):583–667, 2001.

[181] G.W. Swan. *Applications of optimal control theory in biomedicine*, volume 81 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, 1984.

[182] O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quarterly Journal of the Royal Meteorological Society*, 113(478):1311–1328, 1987.

[183] K. Tang, W. Graham, and J. Peraire. Active flow control using a reduced order model and optimum control. AIAA 96-1946, 1996.

[184] S. Tang and N. Aubry. Suppression of vortex shedding inspired by a low dimensional model. *Journal of Fluids and Structures*, 14(4):443–468, 2000.

[185] J.S. Tao, X.Y. Huang, and W.K. Chan. A flow visualization study of feedback control of vortex shedding from a circular cylinder. *Journal of Fluids and Structures*, 10(8):965–970, 1996.

[186] A.C. Taylor, P.A. Newman, L.L. Green, and M.M. Putko. Some advanced concepts in discrete aerodynamic sensitivity analysis. AIAA 2001-2529, 2001.

[187] A.N. Tikhonov and V.Y. Arsenin. *Solutions of ill-posed problems*. Winston, 1977.

[188] T. Tokumaru and P. Dimotakis. Rotary oscillation control of a cylinder wake. *Journal of Fluid Mechanics*, 224:77–90, 1991.

[189] T. Tokumaru and P. Dimotakis. The lift of a cylinder executing rotary motions in a uniform flow. *Journal of Fluid Mechanics*, 255:1–10, 1993.

[190] J.E. Tolsma and P.L. Barton. Hidden discontinuities and parametric sensitivity calculations. *SIAM Journal on Scientific Computation*, 23(6):1861–1874, 2002.

[191] T. Turanyi. Sensitivity analysis of complex kinetic systems.Tools and applications. *Journal of Mathematical Chemistry*, 5(3):203–248, 1990.

[192] S. Ulbrich. On the existence and approximation of solutions for the optimal control of nonlinear hyperbolic conservation laws. In G. Leugering K.-H. Hoffmann and F. Trltzsch, editors, *Optimal control of partial differential equations (Chemnitz 1998)*, volume 133 of *International Series of Numerical Mathematics*, pages 287–299. Birkhauser, 1999.

[193] S. Ulbrich. A sensitivity and adjoint calculus for discontinuous solutions of hyperbolic conservation laws with source terms. Technical Report TR 00-10, Department of Computational and Applied Mathematics, Rice University, To appear in SIAM Journal on Control and Optimization (2002), 2000.

[194] S. Ulbrich. Adjoint-based derivative computations for the optimal control of discontinuous solutions of hyperbolic conservation laws. Technical report, Fakultat fur Mathematik, Technische Universitat Munchen, To appear in Systems and Control Letters (2002), 2002.

[195] S. Ulbrich. Optimal control of nonlinear hyperbolic conservation laws with source terms. Habilitation thesis, Fakultat fur Mathematik, Technische Universitat Munchen, 2002.

[196] S.P. Uryasev. New variable metric algorithms for nondifferentiable optimization problems. *Journal of Optimization Theory and Applications*, 71(2):359–388, 1991.

[197] M. Valorani and A. Dadone. Sensitivity derivatives for non-smooth or noisy objective functions in fluid design problems. In K.W. Morton and M.J. Baines, editors, *Numerical Methods for Fluid Dynamics, vol. 5*, pages 605–631. Oxford Clarendon Press, 1995.

[198] J. Vlcek and L. Luksan. Globally convergent variable metric method for nonconvex nondifferentiable unconstrained minimization. *Journal of Optimization Theory and Aplications.*, 111(2):407–430, 2001.

[199] G. Wahba. A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Annals of Statistics*, 13(4):1378–1402, 1985.

[200] G. Wahba, D.R. Johnson, F. Gao, and J. Gong. Adaptive tuning of numerical weather prediction models: Randomized GCV in 3-dimensional and 4-dimensional data assimilation. *Monthly Weather Review*, 123(11):3358–3369, 1995.

[201] R.W. Walters and L. Huyse. Uncertainty analysis for fluid mechanics with applications. ICASE Report 2002-01, ICASE, 2002.

[202] Y.J. Wang, H. Matsuhisa, and Y. Honda. Loads on lower limb joints and optimal action of muscles for shock reduction. *SME International Journal Series C–Mechanical Systems, Machine Elements and Manufacturing*, 42(3):574–583, 1999.

[203] H. Warui and N. Fujisawa. Feedback control of vortex shedding from a circular cylinder by cross flow cylinder oscillations. *Experiments in Fluids*, 21:49–56, 1996.

[204] P. Wesseling. *Principles of Computational Dynamics*, volume 29 of *Springer Series in Computational Mathematics*. Springer, 2001.

[205] C.H.K Williamson. Vortex dynamics in the cylinder wake. *Annual Review of Fluid Mechanics*, 28:477–539, 1996.

[206] Q. Xu. Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples Part II: Vector formulations and matching conditions. *Journal of Atmospheric Sciences*, 53(8):1123–1155, 1996.

[207] D. You, H. Choi, M. Choi, and S. Kang. Control of flow-induced noise behind a circular cylinder using splitter plates. *AIAA Journal*, 36(11):1961–1967, 1998.

[208] S.T. Zalesak. Fully multidimensional flux-corrected transport algorithm for fluids. *Journal of Computational Physics*, 31:335–362, 1979.

[209] M.M. Zdravkovich. *Flow around circular cylinders. Vol. 1.* Oxford University Press, 1997.

[210] J. Zhang and C. Dalton. A three-dimensional simulation of a steady approach flow past a circular cylinder at low Reynolds number. *International Journal for Numerical Methods in Fluids*, 26:1003–1022, 2000.

[211] S. Zhang, X. Zou, and J.E. Ahlquist. Examination of numerical results from tangent linear and adjoint of discontinuous nonlinear models. *Monthly Weather Review*, 129(11):2791–2804, 2001.

[212] S. Zhang, X. Zou, J.E. Ahlquist, I.M. Navon, and J. Sela. Use of differentiable and nondifferentiable optimization algorithms for variational data assimilation with discontinuous cost functions. *Monthly Weather Review*, 128(12):4031–4044, 2000.

[213] X. Zou, I.M. Navon, M. Berger, P.K. Phua, T. Schlick, and F.X. LeDimet. Numerical experience with limited-memory quasi-Newton methods for large-scale unconstrained nonlinear minimization. *SIAM Journal on Optimization*, 3(3):582–608, 1993.

BIOGRAPHICAL SKETCH

CRISTIAN A. HOMESCU

9/1996– 6/2002 *Florida State University, Tallahassee, FL*
PH.D. Applied Mathematics
Thesis: Optimal control of continuous and discontinuous flow
Advisor: Prof. I.M. Navon

9/1993– 7/1994 *University of Paris Sud, Orsay, France*
M.Sc. in Numerical Analysis

9/1990–9/1993 *University of Craiova, Craiova, Romania*
B.Sc. Mathematics

LIST OF PUBLICATIONS

1. Suppression of vortex shedding for flow around a circular cylinder using optimal control, *International Journal for Numerical Methods in Fluids*,**38**(1):43-69, 2002

2. Numerical and theoretical considerations for sensitivity calculation of discontinuous flow, Accepted for publication in *Systems and Control Letters*, 2002

3. Optimal control of flow with discontinuities, submitted to *Journal of Computational Physics*, 2002

Presentations

1. Optimal control for flow around a rotating cylinder, presented at *Annual Meeting of SIAM*, San Juan, Puerto Rico, July 2000
2. Optimal control of flow with discontinuities (Euler equations), presented at *Fifth Conference of Control and Its Applications of SIAM*, San Diego, July 2001